Equilibrium Neglect and Political Feasibility

Click here for latest version

Bnaya Dreyfuss*

January 31, 2026

Abstract

When voters underappreciate the equilibrium effects of policies, efficient reforms can become politically infeasible. I develop a general framework of *equilibrium neglect*, and use this framework to construct a portable and generally applicable remedy. Using off-path contingent rebates, a social planner can generate a policy that (i) implements the social optimum, (ii) is budget balanced and (iii) is politically feasible even in the presence of equilibrium-neglecting voters. In a survey experiment on congestion pricing with a probability sample from six US metros, I show that respondents are overly pessimistic about traffic-reduction effects of congestion pricing, and that pessimism correlates with strong opposition to the policy. Adding contingent compensation that pays if traffic remains high significantly increases public support, especially among potential compensation recipients. In surveys of civil servants and state legislators, I show how equilibrium neglect affects the policymaking process.

*Harvard University. Email: bdreyfuss@g.harvard.edu. I am grateful to my advisors Raj Chetty, Benjamin Enke, Jesse Shapiro and David Yang for their guidance. I also thank Katherine Coffman, Justin de Benedictis-Kessner, James Druckman, Ed Glaeser, Yannai Gonczarowski, Ori Heffetz, Gabriel Kreindler, David Laibson, Shengwu Li, Matthew Rabin, Alex Rees-Jones, Joshua Schwartzstein, and participants at various Harvard seminars, BEAM 2025, and SITE 2025 for helpful feedback and suggestions. I thank Siena Polls and Benjamin and Joshua Moshes for generously sharing their data, and the Policymakers Lab for administering the civil servants survey. This project is supported by the Chae Family Fund, the Lab for Economic Applications and Policy (LEAP), Kenneth C. Griffin Research Fund, Institute for Quantitative Social Science (IQSS), Institute for Humane Studies (IHS), the Center for American Political Studies (CAPS) and Time-sharing Experiments for the Social Sciences (TESS).

1 Introduction

Economists and the general public often disagree about policies. Despite years of policy advocacy, policies economists typically oppose, such as rent control and tariffs, are viewed positively in the eyes of much of the public, whereas policies economists tend to support, such as carbon taxes and congestion pricing, are viewed negatively (Caplan 2001; Sapienza and Zingales 2013).

There are many legitimate reasons for such disagreements: the models economists use to generate policy recommendations can be misspecified, oversimplified, or ignore aspects that people view as important. However, in many cases, there are good reasons to believe that the way people intuit about the costs and benefits of policies may lead them to the wrong conclusions. In particular, when evaluating a proposed policy, people may underestimate or overlook a crucial element: others' behavioral adjustments in response to it, or its *equilibrium effects* (Dal Bó et al. 2018). This tendency can distort perceptions of a policy's indirect costs or benefits, leading to wrong assessment of its overall impact. I call this bias *equilibrium neglect* (EN). Indeed, there is a large body of work, reviewed below, showing that people often fail to perform contingent reasoning (Niederle and Vespa 2023), specifically when thinking about others (Eyster and Rabin 2005).

Taking this body of work as the starting point, this paper studies the implications for policy design. I ask two questions: (i) descriptively, how does equilibrium neglect constrain policy? (ii) constructively, can policymakers design welfare-improving policies that can win support from equilibrium-neglecting voters? I address these questions using theory and empirics. Theoretically, I first model what it means for agents to neglect equilibrium when evaluating a new policy, and show that such neglect can block efficient reforms. I then use the model to identify a portable and generally applicable remedy. The remedy adds *off-path* contingent rebates to an existing policy, ensuring that if an agent's true benefits from the policy are positive, then the agent's *perceived* benefits from the policy are also positive, even when they neglect equilibrium effects. Empirically, I study these questions in a survey experiment on congestion pricing. Consistent with the theory, I find that respondents' beliefs about the effect of congestion pricing on traffic are overly pessimistic, and that pessimism correlates with opposition to the policy. Moreover, the proposed remedy—a contingent compensation mechanism that pays when congestion remains high—significantly increases support.

I use congestion pricing as a theoretical running example and empirical application. Congestion pricing has a straightforward distinction between direct effects (toll increase) and equilibrium effects (congestion reduction), both of which are unidimensional, easily observed, and precisely measured. Congestion pricing is also notoriously difficult to implement in practice due to political feasibility constraints (Gu et al. 2018), despite consensus among economists and transportation engineers that it can increase welfare.

Finally, there is evidence (reviewed below) suggesting that people underappreciate the congestion-reducing effects of the policy ex ante, but strongly revise their beliefs having observed its effects. These features make congestion pricing a natural application, although the theoretical framework and proposed solution apply broadly to any policy where political feasibility is constrained because behavioral responses are underestimated.

Section 2 describes the contribution of this paper to the literature on behavioral mechanism design, behavioral public policy, failures in contingent reasoning, and the political feasibility of environmental policies.

I present the theoretical framework in Section 3. A social planner chooses a policy from a policy space. Policies shape payoffs and incentives and therefore each policy induces a different game. I model how agents perceive the costs and benefits of a policy change from the status quo, i.e., the *perceived gain* from moving across games within a family of games. The effect of any policy change can be decomposed into a *direct effect*—the effect of a policy change holding others' behavior fixed—and an *equilibrium effect*—the effect of moving to the new equilibrium. I allow agents to underweight or cognitively discount the latter effect, which I call equilibrium neglect. EN can capture different forms of explicit *skepticism*, i.e., believing that others may not change their behavior in response to the policy change, or *naïveté*, i.e., an attenuated perception of equilibrium effects, which are typically complex, less salient and harder to predict than direct effects. Importantly, equilibrium neglect is not a status quo bias. It can generate "grass is greener" cycles where under policy A agents prefer policy B, but under B they prefer A. In other words, while EN means preferences are biased *by* the status quo, they need not be biased *towards* it.

I apply this framework in a road-congestion model and show that equilibrium neglect inflates support for lane expansions, via neglect of induced demand, and reduces support for congestion pricing, via neglect of the toll's effect on congestion. Generalizing these insights, Section 3.4 traces out the implications of EN for political feasibility. The political feasibility of a policy depends on its *perceived gains* (as opposed to real gains) with respect to the status quo through a *feasibility rule* (e.g., a majority rule) that respects standard monotonicity and consensus properties. When agents are sufficiently neglectful, unboundedly beneficial policies can be politically infeasible, and unboundedly harmful policies politically feasible.

Moving beyond mere possibilities, I characterize specific conditions under which EN constrains political feasibility. Intuitively, when high tax burden is associated with large equilibrium benefits, there is a set of agents for whom—under sufficient equilibrium neglect—perceived gains (which emphasize direct costs) are negative, yet actual gains (which also account for equilibrium benefits) are positive. I call such agents a *spurious opposition*. Spurious opposition naturally arises in settings where agents' types determine both equilibrium benefits and their valuation of the taxed action. In the context of road

congestion, agents with high value of time savings are also the ones who value driving the most, which generates a spurious opposition composed of infra-marginal drivers who perceive the cost of the toll, but neglect its equilibrium benefits from shorter commute times.

Turning to the constructive question, Section 3.5 describes my proposed remedy. The logic applies generically but is easiest to see in the context of congestion pricing. Agents spuriously oppose a toll when they underweight its equilibrium effect relative to the direct effect. The proposed remedy augments the toll with contingent compensation paid when congestion remains high. A sufficiently high contingent payment makes the direct effect positive and closes the gap between perceived and actual gains; thus, if the original toll is politically feasible absent equilibrium neglect, the augmented policy is politically feasible for *any* degree of equilibrium neglect. I call such a policy *EN-robust*. The contingent payment does not create perverse incentives: because compensation depends only on aggregate congestion and not on any individual's action, individual incentives, i.e., the cost of driving, are unchanged. Thus, the equilibrium under the EN-robust toll coincides with that under the original toll; the policies are *outcome equivalent*. And because compensation is triggered only when congestion remains high, i.e., off the equilibrium path, no contingent payment is made in equilibrium, so the original and EN-robust tolls are *deficit equivalent*. Theorem 1 generalizes this intuition to any setting with quasilinear utilities and transfers. It shows that starting from any status quo, any policy a social planner wishes to implement has an EN-robust counterpart that is outcome- and deficit-equivalent.

While contingent compensations maintain the deficit in equilibrium, they may generate off-path deficits. Voters and policymakers may be skeptical of such off-path promises. Can we construct contingent compensations such that the deficit is constrained even when payment is triggered? Theorem 2 shows that the answer is yes, as long as the compensation can be conditioned on an individual agent's past behavior. In this case, we can ensure that the deficit of the EN-robust policy never exceeds the status-quo deficit, even off the equilibrium path. The intuition combines two key insights: first, compensation is triggered when agents do not respond to the policy, which generates revenue surplus that can be used to fund the payment. For instance, in congestion pricing, if congestion remains high despite the toll, toll revenue is also high. Second, a tailored compensation based on agents' status-quo actions can target potential members of the spurious opposition. In congestion pricing, compensation need only be directed to those who drive in the status quo, as they are the ones directly affected by the toll. Combining these two insights, I show that conditioning on past behavior, there is always a contingent compensation that does not increase the deficit relative to the status quo.

Empirically, I study equilibrium neglect and my proposed remedy in the context of congestion pricing, focusing on stated opinion in polls. Stated opinion is a direct measure

of the relevant determinant of political feasibility: policies that poll badly are unlikely to get implemented (Butler 2011). Later, I turn to survey policymakers directly.

As motivation, in Section 4.1 I analyze existing polls from seven cities that implemented congestion pricing, showing that support consistently increases post-implementation. I then focus on New York City—the most recent of the seven to implement congestion pricing—for which I have individual response data: two Siena polls conducted in New York state before (December 2024, N = 834) and after (March 2025, N = 806) the policy's implementation. I classify respondents by the Manhattan commuting intensity of their residential location, derived from ACS traffic flows. Support for congestion pricing increases substantially more among respondents who live in commuting-intense areas (29.8 net support points; SE = 8.48), consistent with equilibrium neglect generating spurious opposition that, after implementation, dissipates in geographical areas where benefits are visible.

To study equilibrium neglect more directly, I turn to my own survey experiment, described in Section 4.2. The survey has two goals, corresponding to the two research questions: descriptive and constructive. First, I document how equilibrium neglect—in the form of pessimistic beliefs about traffic reduction—constrains support for congestion pricing. Second, I test the effectiveness of the remedy: respondents offered contingent compensation—triggered when traffic remains high—should be more supportive of the policy. Establishing equilibrium neglect does not mechanically imply the remedy will work: if agents neglect equilibrium effects due to cognitive constraints, they may similarly struggle to understand the contingent nature of the remedy, potentially undermining its effectiveness.

Respondents are recruited from AmeriSpeak, a probability-based panel by NORC at the University of Chicago. The sample consists of 2,056 adults residing in one of six major U.S. metropolitan areas: Boston, Chicago, Dallas, Los Angeles, San Francisco, and Washington D.C. The survey uses a two-stage, between-subject design. In the first stage, all respondents see a standard congestion pricing policy modeled after the originally proposed $15 toll under NYC's Central Business District Tolling Program. Respondents then report their beliefs about congestion reduction, and their commute habits.

Respondents overwhelmingly oppose standard congestion pricing. The share of respondents opposing the policy is 63.2%, while 20.7% support it, corresponding to -42.5 points *net support*. Respondents also substantially underestimate the benefits from congestion pricing. Half the respondents are asked to predict changes to their own driving times. Around 39% indicate they expect no change, another 20% that traffic conditions will *worsen*, and 41% anticipate some improvement. The other half estimate the effect of NYC's congestion pricing on travel times through the Holland Tunnel, which provides a concrete, easily verifiable benchmark with a known ground truth. I find similarly high levels of underappreciation: 65% of respondents underestimate the actual 12-minute

reduction, and 33% estimate little to no improvement from the 23-minute baseline.

This pessimism correlates with opposition to the policy: net support among those who predict any amount of time savings in their commute is 36.7 points higher than those who predict no savings (-62.4 vs. -25.7; $p < 0.001$). Beliefs about time savings explain 5 percent of the variation in support in the sample, but this increases to 19 percent among respondents with above-median confidence in their beliefs, and up to 38 percent among respondents who drive to downtown more frequently, and those with above-median value of time (proxied by income)—i.e., those most affected by congestion, and who value time savings most. Translating beliefs into dollar units, I combine beliefs about traffic with value of time (inferred from income) to construct a perceived benefit index for each respondent. An additional dollar of perceived benefit corresponds to a 2-point increase in net support (SE = 0.33). Together, these findings suggest prevalent underappreciation of equilibrium effects that constrains support, consistent with spurious opposition.

I next turn to the constructive question: can the proposed remedy increase support? In the second stage, respondents are asked to state their support for one of three randomly assigned modified policies. The *EN-robust* treatment adds a "Monitoring & Assurance" component: respondents are told that traffic will be monitored and data will be publicly available, and that if after the implementation traffic decreases by less than 8% in a given month (compared to that month's baseline level), workers who commute to work in the downtown area will receive a \$120 monthly compensation.[1] The *Control* treatment includes only the monitoring component, without compensation. The third treatment, *Toll-Reduction*, lowers the toll from \$15 to \$9 and serves as a benchmark to measure the impact of EN-robustness against a more conventional adjustment margin.

I find a significant effect of contingent compensation on support, concentrated among those it targets. Among respondents who work in the downtown area ($N = 184$)—those eligible for compensation if congestion remains high—net support in the EN-robust condition increases from baseline by 48 points more than in Control (SE = 12.21). The treatment effect of the contingent compensation on the whole sample ($N = 2,056$) is a more moderate yet highly statistically significant 11 point increase (SE = 3.85). These results suggest that contingent compensation increases support among the group it targets. By alleviating the concerns of those who are negatively affected, it can be an effective tool for increasing policy support. In contrast, the Toll-Reduction condition shows no meaningful improvement relative to Control.

Finally, to understand how these dynamics translate into political feasibility, Section 5 turns to the supply side of policymaking, surveying two key groups of policymakers: state legislators and civil servants. I elicit policymakers' own beliefs and attitudes about congestion pricing—both the standard and EN-robust versions—as well as their second-

1. This policy implements the off-path budget balanced contingent compensation, using workplace location as a proxy for status-quo driving frequency.

order beliefs, i.e., their perceptions of the public's beliefs and attitudes. State legislators exhibit equilibrium neglect at rates similar to the general public, and as with the public, these beliefs strongly predict their policy support. Consistent with that, state legislators' own support for the EN-robust version of congestion pricing increases substantially, as does their estimate of public support for it. In contrast, civil servants largely demonstrate a textbook understanding of the policy's effects, and support standard congestion pricing. However, their support for EN-robust congestion pricing is *lower* than their support for the standard version, even as they correctly predict it will be more popular with the public. These findings highlight a potential disconnect, where the experts who design policies may disapprove of the very remedies that could make those policies politically viable with elected officials and the public they are meant to serve.

2 Related Literature

My main contribution is to construct and test a theory-driven practical remedy for a feasibility challenge driven by cognitive limitations, using off-path contingent payments.[2] The core insight contributes to behavioral mechanism design, and its practical implications contribute to the literature on behavioral public economics.

I start with the former. The paper adds to a growing literature on how cognitive limitations shape the design of mechanisms. The strategic simplicity program (Li 2017; Bó and Hakimov 2023; Pycia and Troyan 2023; Nagel and Saitto 2025) characterizes strategy-proof mechanisms by the amount of contingent or forward-looking reasoning agents need in order to identify their dominant strategy. A related strand studies which social outcomes and rules a planner can implement with cognitively limited participants: de Clippel (2014) and de Clippel et al. (2019) provide implementation results when choice correspondences may depart from the classical benchmark and under level-k equilibrium, respectively. Gavan and Penta (2025) extend Nash implementation to admit deviations from equilibrium behavior while restricting outcomes to an acceptable set. Closest in spirit to my remedy, but different in context, Gonczarowski et al. (2023) and Dreyfuss et al. (2025), both modify existing matching mechanisms in outcome-equivalent ways to handle departures from the classical model. The key distinction between this body of work and my paper is that rather than studying incentive compatibility, i.e., behavior

2. Mechanisms that use off-path payments contingent on others' behavior ("side bets") have already appeared in the literature. For example, Myerson (1981) shows that with correlated valuations, using side bets can extract the full surplus, later expanded in Crémer and McLean (1988). In platform design, they arise as "insulating tariffs" (Weyl 2010), allowing platforms to effectively select among multiple equilibria and avoid coordination failures. In these papers, off-path payments primarily serve as an analytical device—for example, to clarify the role of independence assumptions or to abstract from equilibrium selection and focus on allocation choice. In contrast, I treat off-path contingent payments as a practical design tool that can restore the feasibility of otherwise desirable mechanisms when opposition is driven by biased beliefs.

within a mechanism, I study how cognitive limitations affect individual rationality, i.e., preferences *over mechanisms*, which has received less attention.

Moving from abstract mechanisms to applications, this paper studies practical policy design questions in the presence of cognitively limited or biased agents. In doing so it joins the behavioral public economics literature, reviewed in Bernheim and Taubinsky (2018). For example, Chetty et al. (2009), Rees-Jones and Taubinsky (2020), and Allcott et al. (2019) document systematic biases and analyze their potential implications for the design of welfare-maximizing taxation. A similar distinction applies here: much of this work analyzes the *behavior* of cognitively limited agents under different policies, whereas I study how these limitations constrain *adoption*, and the implications for policy design.

This paper also contributes to the literature on the economic implications of failures in contingent thinking, reviewed in Niederle and Vespa (2023). A central failure is misunderstanding how actions relate to incentives, which matters both theoretically and empirically (e.g., Eyster and Rabin 2005; Cohen and Li 2025; Esponda and Vespa 2014). Most relevant here is Dal Bó et al. (2018), who document systematic underappreciation of equilibrium effects in the lab, showing that participants fail to recognize that the other player will adjust their behavior in response to incentive changes, leading them to favor a socially dominated game. I take the core insight of this paper as my starting point: I generalize their framework to an abstract policy setting, derive implications for political feasibility, and use it to construct a remedy. Empirically, I document equilibrium underappreciation in a high-stakes policy setting and test the remedy in that same setting. Other relevant papers include Elmendorf et al. (2025), who document systematic underappreciation of housing price responses to supply shocks, and Micheletto et al. (2025), who show that when agents overestimate their own demand elasticity relative to others, they may support an "inverse Ramsey" rule, placing higher taxes on more elastic goods. More broadly, the paper relates to the growing literature on people's misspecified subjective models of the economy (Spiegler 2020; Stantcheva 2022; Chinoy et al. 2023; Binetti et al. 2024).

Finally, I contribute to the large literature on public support for environmental policies, reviewed in Carattini, Carvalho, et al. (2018), and for congestion pricing in particular, reviewed in Gu et al. (2018). This literature consistently finds that support for congestion pricing and other environmental policies is higher after implementation, and that people are often skeptical of effectiveness beforehand but typically revise beliefs and attitudes after experiencing the effects (see, for example, Schuitema et al. (2010) and Andersson and Nässén (2016) in the context of congestion pricing, Carattini, Baranzini, et al. (2018) in the context of a garbage tax, and Murray and Rivers (2015) in the context of carbon taxes). Both findings are consistent with equilibrium neglect that dissipates once people are exposed to the true benefits of the policy. My first contribution to this literature is to offer a general framework that can explain and organize these findings.

My second and main contribution is to propose a potential remedy to common adoption frictions discussed in the literature. Comparing my proposed remedy to other frequently discussed tools, despite the fact that perceived effectiveness predicts support, evidence on the persistent effects of information-provision treatments is mixed.[3] In contrast, real changes to the actual content of a policy have been shown to shape support. In particular, reducing the size of the tax (Baranzini et al. 2021) and earmarking revenues for specific purposes (Jaensirisak et al. 2005) are generally associated with higher support for congestion pricing. My proposal combines these approaches: it addresses over-pessimism not by attempting to change it directly, but by leveraging it into a policy change that increases perceived benefits, especially for over-pessimistic voters. The contingent compensation mechanism can be thought of as a specific form of earmarking: explicit revenue recycling that pays more when outcomes are worse, potentially persuading those most concerned about them.

3 Theory

3.1 Preliminaries

There is a set N of agents, where N is a continuum endowed with a measure space structure $(N, \mathcal{N}, \mu)$.[4] Throughout, I assume that the relevant functions are measurable. Each agent $n \in N$ has a set of actions A_n. The set of action profiles is $A = \prod_{n \in N} A_n$.

A *policy profile* $T \in \mathbb{T}$ is a collection of agent-specific policies, $T = (T_n)_{n \in N}$. When context is clear, I refer to a policy profile simply as a *policy*. Agents have preferences over action profiles and their own policy component, represented by a bounded utility function $v_n : A \times \mathbb{T}_n \to \mathbb{R}$, denoted $v_n(a; T_n)$, where $\mathbb{T}_n$ is the set of possible policies for agent n. A profile of utility functions is denoted $v = (v_n)_{n \in N}$.

It is useful to view policies as operating through *consequences*. In this representation, each agent n has a consequence space C_n, and a policy is a function $T_n : A \to C_n$ mapping action profiles to consequences. Indeed, with a sufficiently rich consequence space, there exists $u_n : A \times C_n \to \mathbb{R}$ such that

$$v_n(a; T_n) = u_n(a, T_n(a)) \quad \text{for all } a \in A \text{ and } T_n \in \mathbb{T}_n.$$

Consequences can capture transfers, public goods, infrastructure, or resource allocations; policies specify how actions map to these (monetary or non-monetary) outcomes. A

3. For example, Douenne and Fabre (2022) show small effects for carbon taxes; Baranzini et al. (2021) show mixed results for congestion pricing, and Allcott and Knittel (2019) show null results in the context of fuel economy; however, see for example Semken (2025) and Dechezleprêtre et al. (2025) who find positive effects in the context of climate change.

4. The proofs of the main results are extended for the case of finite N.

policy T_n is *unilateral* if there exists a consequences representation where consequences depend only on n's own behavior, i.e., $T_n(a_n, a_{-n})$ is independent of a_{-n}. Otherwise, the policy is *multilateral*.

A social planner first chooses a policy profile T, and then each agent chooses an action $a_n \in A_n$. Each policy profile thus induces a normal-form game.[5] An *economy* $\mathcal{E} = \langle N, A, v, \mathbb{T} \rangle$ is a policy-parametrized family of normal-form games, and a policy change from T to T' is a transition between games within the family. I assume that for each policy profile $T \in \mathbb{T}$, a unique Nash equilibrium is selected. For simplicity, I further assume that it is pure, as is the case in all congestion games and, more generally, all potential games (under regularity conditions). The Nash equilibrium under policy profile T is denoted by $a^*(T)$. Agent n's best response to a profile a_{-n} and a policy T_n is denoted $a_n^*(a_{-n}, T_n)$. (Appendix B extends the framework and allows for aggregate noise that induces a distribution over equilibrium outcomes. It shows that equilibrium neglect and its remedies naturally extend to this setting. I discuss this extension at relevant points throughout this section.)

To illustrate the different results in this section, I develop a simple congestion model.

Example (Road Congestion and Transportation Policies: Setup.) Consider the following congestion model. A transportation network connects where people live and where they work. In all but one application, I abstract away from route choice and consider a single road. There is a unit continuum of commuters N. Each commuter $n \in N$ makes a binary choice between driving ($a_n = 1$) and taking public transit (PT; $a_n = 0$). The policy consists of the level of infrastructure I, a toll p drivers must pay, and a lump-sum transfer k. Utilities are normalized so that the utility from taking public transit equals the lump-sum transfer k. Let $\alpha(a)$ denote the share of commuters who drive in profile a. Agent n's utility from driving is given by

$$v_n\left((1, a_{-n}); T\right) = \theta_n Q(\alpha(a); I) - (b + p) + k,$$

where $\theta_n \geq 0$ is agent n's value of time, and the function $Q(\alpha(a); I)$ is time saving from driving given a level of congestion α and a level of infrastructure investment I; it is monotonically increasing in I, monotonically decreasing and continuously differentiable in α and $Q(0; I) > 0$ for all I, meaning that when there is no congestion, driving is faster than public transit. The parameter $b > 0$ captures driving costs (e.g., gas, depreciation) net of public transit ticket price. The value of time θ is distributed according to a CDF F which is continuously differentiable and supported over $[0, \theta_{max}]$. Throughout, I assume that θ_{max} is sufficiently large so that a positive measure of commuters choose to drive.

5. For any policy profile $T \in \mathbb{T}$, the induced game is $G(T) \equiv \left\langle N, (A_n)_{n \in N}, (v_n(\cdot; T_n))_{n \in N} \right\rangle$.

The policy space is $\mathbb{T} \subset \mathbb{R}^3$, representing the level of investment, toll and transfers.[6] The *equilibrium deficit* of a policy T, denoted D_T^*, is given by

$$D_T^* = k + I - p \cdot \alpha(a^*(T)),$$

i.e., government spending on transfers and infrastructure minus (equilibrium) toll revenue. Policies are constrained so that their deficit does not exceed a threshold $\overline{D}$.[7] Formally:

$$\mathbb{T} = \left\{ T \in \mathbb{R}^3 : D_T^* \leq \overline{D} \right\}.$$

The unique equilibrium in this game is a cutoff equilibrium, where only agents whose value of time is above some threshold choose to drive:

Result 1. *In the congestion model, for any $T \in \mathbb{T}$ there is a unique Nash equilibrium $a^*(T)$ characterized by the cutoff $\theta^*(T)$ where $a_n = 1$ if and only if $\theta_n \geq \theta^*(T)$.*

All proofs are in the Appendix.

3.2 Equilibrium Neglect

To understand how equilibrium neglect shapes political feasibility, we first need to model how agents evaluate policy changes. A policy change affects agents through two channels: first, directly, by changing their payoffs and second, indirectly, through its effect on others' behavior. The key intuition is that agents assessing a proposed policy may fail to fully account for the latter.

Consider a status-quo policy T^0 with the corresponding equilibrium $a^0 = a^*(T^0)$. When a new policy T is proposed, the *total gain* for agent n is given by

$$\delta_n(T, T^0) = v_n\left(a^*(T); T_n\right) - v_n\left(a^*(T^0); T_n^0\right),$$

I often refer to the total gain as the *actual* gain. A profile of total gains from T given T^0 is denoted $\delta(T, T^0) = \left(\delta_n(T, T^0)\right)_{n \in N}$, and a generic profile of total gains is denoted $\delta \in \mathbb{R}^N$.

Let $\hat{a}_n = a_n^*\left(a_{-n}^0, T_n\right)$ denote agent n's best response to policy T, fixing others' actions at the status quo. Any total gain can be decomposed into *direct* and *equilibrium* effects:

$$\delta_n(T, T^0) = \underbrace{v_n\left((\hat{a}_n,\, a_{-n}^0); T_n\right) - v_n\left(a^0; T_n^0\right)}_{\text{direct effect}} + \underbrace{v_n\left(a^*(T); T_n\right) - v_n\left((\hat{a}_n,\, a_{-n}^0); T_n\right)}_{\text{equilibrium effect}}. \quad (1)$$

6. In this case, a policy T_n can be modeled as operating through consequences if the consequence space contains time saving and wealth, i.e., $C_n \subseteq \mathbb{R}^2$, where $c = (Q, I)$ represents time savings Q and wealth W and $u_n(a, c) = \theta_n Q + W$.

7. If the government needs to raise a minimal revenue, we let $\overline{D} < 0$.

The decomposition is obtained by adding and subtracting the counterfactual object $v_n\left((\hat{a}_n,\, a^0_{-n}); T_n\right)$, which is the agent's utility under the proposed policy T, with others' behavior held at the status quo. The direct effect captures the impact of the policy change from T^0 to T absent any equilibrium adjustments (i.e., fixing others' actions at a^0_{-n}). The equilibrium effect captures how others' behavioral responses to the policy change affect the agent's welfare, holding the policy fixed at T.

Agents suffer from *equilibrium neglect* when their *perceived* gain from a proposed policy systematically underweights its equilibrium effects:

Definition 1 (Perceived Gain)**.** *Let $\eta_n \in [0,1]$. Agent n's **perceived gain** from proposed policy T under the status quo T^0 is given by:*

$$\tilde{\delta}_n(T, T^0; \eta_n) = v_n\left((\hat{a}_n,\, a^0_{-n}); T_n\right) - v_n\left(a^0; T^0_n\right) + \eta_n\left(v_n\left(a^*(T); T_n\right) - v_n\left((\hat{a}_n,\, a^0_{-n}); T_n\right)\right).$$

The *perceived* gain $\tilde{\delta}_n$ deviates from the real gain δ_n as a function of equilibrium neglect, parameterized by η. I index η_n by agent n but often generically refer to the parameter as η. A profile of equilibrium-neglect parameters is denoted $\boldsymbol{\eta} = (\eta_n)_{n\in N}$. For any T, T', denote the profile of perceived gains by $\tilde{\delta}(T, T'; \boldsymbol{\eta}) = \left(\tilde{\delta}_n(T, T'; \eta_n)\right)_{n\in N}$, and a generic profile of perceived gains by $\tilde{\delta}$. The following definition quantifies the potential distortion caused by equilibrium neglect:

Definition 2 (Spurious Opposition and Coalition)**.** *Given profiles of actual and perceived gains, δ and $\tilde{\delta}$, the **spurious opposition** is the set of all agents with weakly positive actual gains, and negative perceived gains:*

$$\left\{n \in N : \delta_n \geq 0 \;\; \textit{and} \;\; \tilde{\delta}_n < 0\right\}.$$

*Similarly, a **spurious coalition** consists of agents with negative actual gains but weakly positive perceived gains.*

3.2.1 Discussion

The parameter η can capture different psychological mechanisms leading to equilibrium neglect. These fall into two general categories: *equilibrium naïveté* and *equilibrium skepticism*. Equilibrium naïveté arises when agents partly overlook equilibrium effects due to its complexity. Direct effects are typically straightforward to understand and compute.[8] In contrast, fixed-point reasoning is difficult and often requires perceiving and aggregating over a population distribution, which adds another layer of complexity. Such difficult-to-calculate objects may be under-appreciated and discounted by η. Alternatively, the

8. Notice that the agent's own optimization is included in the direct effect. The model and the main results could be extended to accommodate agents neglecting their own optimization. Doing so would significantly complicate notation, and I speculate will not add much insight to the core results.

perceived gain, $\tilde{\delta}_n$ could be interpreted probabilistically, in which case $1-\eta$ is the probability that the agent overlooks the equilibrium effect when evaluating a proposed policy. A related interpretation is that η captures noisy cognition (Enke and Graeber 2023): agents predict the total gain by relying on "cognitive signals" generated through sequential cognitive sampling or deliberation process, which attenuates the more complex equilibrium component towards zero.[9]

Equilibrium skepticism reflects a more explicit belief about the equilibrium. For example, η can capture the perceived probability of others' responsiveness: agents may assign a non-zero probability $(1-\eta)$ to the event that others' behavior will remain fixed at the status quo, thereby neglecting others' elasticity. This may reflect skepticism about others' willingness or ability to adjust behavior in response to incentives, or may arise from attributing others' current behavior to fixed dispositions rather than to their response to a given situation, resulting in systematic under-appreciation of others' elasticity.

While naïveté and skepticism are psychologically distinct—in one agents overlook an important input, while in the other they hold a misspecified model—none of my results hinge on the particular psychological mechanism driving equilibrium neglect. In what follows I remain largely agnostic to the interpretation of η.

The formulation has the standard structure nesting the classical model as a special case: when $\eta = 1$, the *perfect foresight* case, equilibrium effects are correctly perceived ($\tilde{\delta}_n(T, T^0; 1) = \delta_n(T, T^0)$). The other extreme is *full neglect*, $\eta = 0$, where agents completely disregard changes in others' behavior and only perceive the direct effect. The intermediate range $\eta \in (0, 1)$ captures *partial neglect*.[10]

Aggregate Uncertainty Appendix B extends equilibrium neglect to an environment with aggregate uncertainty. In the extension, after a policy is chosen, a common shock realizes, and each policy induces a distribution over equilibrium outcomes rather than a single outcome. Equilibrium-neglecting agents understate how much a policy change shifts this distribution: they evaluate a new policy using a perceived mixture of the status-quo equilibrium distribution and the true distribution induced by the new policy. Beyond the interpretations discussed above, the smoothness of mixing distributions allows equilibrium neglect in this setting to naturally capture underestimation of the population elasticity to incentives by attenuating the perceived shift in the aggregate distribution.

9. The expression in Definition 1 is the mean posterior of this process if (i) predicting the (typically simple) direct effect involves minimal noise and (ii) predicting the (typically complex) equilibrium effect involves a non-trivial amount of cognitive noise, and the prior is centered around zero (a natural cognitive default if, e.g., equilibrium effects average zero).

10. The framework can also capture *over*-weighting of equilibrium effects when $\eta > 1$. While this form of bias can be important in policy discourse (and is perhaps common even, or maybe especially, among economists), I leave it for future work.

Example (Over-investment in roads.) Recall the congestion model from page 9. Let the status-quo policy be $T^0 = (0, 0, 0)$ and denote the share of equilibrium status-quo drivers $\alpha^0 = \alpha(a^0)$. Consider a road-investment policy funded by a lump-sum tax: $T = (h, 0, -h)$. Focusing on the marginal dollar (i.e., small values of h), for agents currently taking public transit (i.e., agents with $\theta_n < \theta^*(T^0)$), the actual gain is the perceived gain: a unit increase in investment is simply a tax increase. Therefore, $\tilde{\delta}_n(T, T^0; \eta) = -h$. The perceived gain of those who currently drive ($\theta_n \geq \theta^*(T^0)$), is given by

$$\tilde{\delta}_n(T, T^0; \eta_n) = h \cdot \left(\underbrace{-1 + \theta_n \frac{\Delta Q(\alpha(T^0); 0)}{\Delta I}}_{\text{direct effect}} + \eta_n \underbrace{\theta_n \frac{\partial Q(\alpha^0; 0)}{\partial \alpha} \cdot \frac{\Delta \alpha(a^*(T^0))}{\Delta I}}_{\text{equilibrium effect}} \right).$$

The first part is the benefit from shorter commute time when infrastructure improves, and the second part captures the equilibrium increase in road usage, which partly (or even fully) offsets it (Duranton and Turner 2011). Since the equilibrium effect (slower commute) is negative, agents with $\eta < 1$ over-value the benefits from road investments. For example, fixing $\eta < 1$ for all agents, there is always a non-empty set of agents who perceive a positive gain ($\tilde{\delta}_n(T, T^0; \eta) > 0$) but whose real gain is negative ($\delta_n(T, T^0) < 0$), a spurious coalition, creating a gap between the normative and the perceived willingness to accept a tax increase, potentially leading to over-investment in roads.[11]

3.3 Political Feasibility

How do perceived gains translate to political feasibility?

Definition 3 (Political Feasibility Rule)**.** *A **political feasibility rule** is a mapping $f : \mathbb{R}^N \to \{0, 1\}$. Given a status-quo policy T^0 and a proposed policy T with associated profile of perceived gains $\tilde{\delta}(T, T^0) \in \mathbb{R}^N$, policy T is **politically feasible** under the status quo T^0 with respect to the feasibility rule f if $f(\tilde{\delta}) = 1$.*

I impose two restrictions on the political feasibility rule:

Assumption 1 (Monotonicity and Consensus)**.** *The political feasibility rule f satisfies the following properties:*

1. *Monotonicity: if $f(\tilde{\delta}') = 1$ and $\tilde{\delta}_n \geq \tilde{\delta}'_n$ for all n, then $f(\tilde{\delta}) = 1$.*

11. This idea is captured by Goodwin (1997): "From the late 1950s onwards the transport planning orthodoxy was 'predict and provide'... first we forecast how much traffic there will be, and then we build enough road space to accommodate it. [...] 'Predict-and-provide' actually meant, inevitably, 'predict-and-*under*provide' [since] supply of road space will not—because it cannot—be increased to match demand. [Eventually, a transport committee] reported that road construction in conditions of congestion normally results in an increase in the total volume of traffic, hence a shorter period of relief from congestion, thereby consigning to history the untenable assumption that the total volume of traffic is unaffected by traveling conditions."

2. *Consensus: if* $\tilde{\delta}_n < 0$ *for all* $n \in N$ *then* $f(\tilde{\delta}) = 0$. *Conversely, if* $\tilde{\delta}_n \geq 0$ *for all* $n \in N$ *then* $f(\tilde{\delta}) = 1$.

A common and intuitive political feasibility rule is a minimal required share of weighted support:

$$f(\tilde{\delta}) = 1 \iff \int_{n \in N} \lambda_n \mathbf{1}\{\tilde{\delta}_n \geq 0\} d\mu(n) = S_\lambda(\tilde{\delta}) \geq \beta,$$

where λ_n is agent n's (weakly positive) influence weight with $\int_{n \in N} \lambda_n d\mu(n) = 1$, and $\beta \in (0, 1]$ is a feasibility threshold. This rule requires that a policy T has sufficient (influence-weighted) support to overcome any opposition. Two focal cases are $\beta = 1$, where only Pareto improvements (among individuals with positive weight) are politically feasible (e.g., because a vocal minority can maintain the status quo), and $\beta = \frac{1}{2}$ with equal influence weights—the simple majority case.

Alternatively, a rule might implement a weighted Kaldor–Hicks test, requiring that the weighted sum of welfare changes across all individuals exceeds some threshold:

$$\int_{n \in N} \lambda_n \tilde{\delta}_n d\mu(n) > B.$$

A natural choice for the minimum required aggregate welfare improvement is $B = 0$, with $B > 0$ capturing any standard frictions or status-quo bias.

3.4 Implications

Having defined the two main objects of this paper—equilibrium neglect and political feasibility—we want to understand the degree to which the former imposes a constraint on the latter.

3.4.1 Cyclical Preferences and Unbounded Reversals

I start by looking at the implications of EN for individual preferences which political feasibility depends on. The first question one might ask is whether an equilibrium-neglecting agent has coherent policy preferences. Proposition 1 shows that the answer is generally negative:

Proposition 1 (Cyclical Preferences)**.** *Let* $\succ_n$ *be a preference relation over the space of policies* $\mathbb{T}$*, defined as* $T \succ_n T' \iff \tilde{\delta}_n(T, T'; \eta_n) > 0$*. Then for any* $\eta_n < 1$*, there exists an economy* $\mathcal{E}$ *such that* $\succ_n$ *is not asymmetric, i.e.,* $T \succ_n T' \succ_n T$ *for some* $T, T' \in \mathbb{T}$*.*

Note that asymmetry violations are more severe than transitivity violations: they admit *pairwise* contradictions (not just chains). The proof can be shown in the context of the running congestion example.

Example (Cyclical transportation preferences.) Consider the case where *two* routes, R_1 and R_2, connect where people live and work. Commuters choose between public transit ($a_n = 0$) and driving on route R_i ($a_n = i$, $i \in \{1, 2\}$). Time savings from driving on R_i is

$$Q_i(\alpha_i; T) = s_i - \alpha_i(a),$$

where $\alpha_i(a)$ is the share of commuters choosing R_i in profile a, and s_i is chosen by the planner. In particular, let $\mathbb{T} = \{T_1, T_2\} = \{(s_1, s_2) \in \{(\overline{s}, \underline{s}), (\underline{s}, \overline{s})\}\}$, interpreted as choosing which route gets traffic management (e.g., traffic light) priority, where T_i prioritizes route R_i. Equilibrium time savings on both routes must be equal, i.e.,

$$s_1 - \alpha_1(a) = s_2 - \alpha_2(a),^{12}$$

implying actual gains from a policy change are always zero: $\delta_n(T_i, T_{-i}) = 0$ for $i \in \{1, 2\}$ and all $n \in N$. However, perceived gains for drivers are given by:

$$\tilde{\delta}_n(T_i, T_{-i}; \eta_n) = \theta_n(\overline{s} - \underline{s}) + \eta_n \theta_n(\alpha_i(a^*(T_{-i})) - \alpha_i(a^*(T_i))) = \theta_n(1 - \eta_n)(\overline{s} - \underline{s}),$$

where the second equality follows from the equilibrium condition. The direct gain from T_i under T_{-i} is the time savings increase $(\overline{s} - \underline{s})$ on the newly prioritized R_i. The equilibrium effect completely offsets the direct gain as drivers re-optimize and congestion on R_i increases until gains are completely competed away: $-(\overline{s} - \underline{s})$. Since $\overline{s} > \underline{s}$, for any $\eta_n < 1$, perceived gains are strictly positive: under T_1, agents perceive benefits from T_2, and under T_2, they perceive gains from T_1, generating a 2-cycle. Intuitively, under T_1, agents look at R_2 and see an emptier road. Neglecting equilibrium, they believe that giving this emptier road traffic-management priority will make it faster for them, neglecting the increase in drivers that will follow.

As the example highlights, EN is not a status-quo bias: here, EN generates a "grass is greener" cycle, making a change appear beneficial from the perspective of the current equilibrium. Therefore, while under EN policy preferences are biased *by* the status quo, they are not necessarily biased *towards* it.

Another question that might arise is whether EN-driven distortions on political feasibility are bounded. The following proposition shows that the answer is generally negative: with sufficient neglect, agents may oppose an arbitrarily beneficial policy, and may support an arbitrarily harmful one.

Proposition 2 (Unbounded Distortions)**.** *Fix any feasibility rule f satisfying monotonicity and consensus, and any $K > 0$. There exists an economy $\mathcal{E}$ and $\boldsymbol{\eta} \in \mathbb{R}^N$ where:*

1. (Infeasible unbounded gains) There exist $T^0, T^+ \in \mathbb{T}$ such that $\delta_n(T^+, T^0) \geq K$ for

12. To maintain equilibrium uniqueness, we can impose any arbitrary selection criterion for which drivers choose which route.

all $n \in N$*, yet* $f(\tilde{\delta}(T^+, T^0; \boldsymbol{\eta})) = 0$.

2. *(Feasible unbounded losses) There exist* $T^0, T^- \in \mathbb{T}$ *such that* $\delta_n(T^-, T^0) \leq -K$ *for all* $n \in N$*, yet* $f(\tilde{\delta}(T^-, T^0; \boldsymbol{\eta})) = 1$.

3.4.2 Spurious Opposition

Propositions 1 and 2 show the existence of settings where EN distorts personal and societal policy rankings. How relevant are these to real-world economic settings? For EN to impose a constraint on feasibility, there needs to be a spurious opposition: a share of agents whose actual gains are positive, but whose perceived gains are negative. The following proposition characterizes conditions under which achieving a social outcome with budget-balanced taxation generates spurious opposition.

From this point on, I focus on a particular class of economies. In a *quasi-linear economy*, the policy space is the set of transfer schemes $\mathbb{T} \subseteq (A \to \mathbb{R})^N$, and utilities are quasi-linear in transfers:

$$v_n(a; T_n) = w_n(a) + T_n(a) \text{ for all } n \in N.$$

A particularly important special case is an *aggregative economy*, in which there exists an *aggregator* that maps action profiles to a real aggregate quantity: $\alpha : A \to \mathcal{Q} \subset \mathbb{R}$, where $\alpha(a) = \alpha(a')$ whenever $a = a'$ a.e. (i.e., $\mu(\{n : a_n \neq a'_n\}) = 0$). Agents' utility depends on their own action a_n and the aggregate $\alpha(a)$: $\nu_n : A_n \times \mathcal{Q} \to \mathbb{R}$, so that $w_n(a) = \nu_n(a_n, \alpha(a))$ for all $n \in N$.

A policy's *deficit at profile* a is defined as $D_T(a) = \int_{n \in N} T_n(a) d\mu(n)$, and the *equilibrium deficit* of policy T is $D^*_T = D_T(a^*(T)) = \int_{n \in N} T_n(a^*(T)) d\mu(n)$.

I characterize conditions for a spurious opposition to arise in the following setting: let $\mathcal{E}$ be an aggregative economy, where each agent $n \in N$ has type θ_n distributed according to CDF F, continuously differentiable with strictly positive density on $[0, \theta_{max}]$. Utility depends on agents' types:

$$\nu_n(a_n, \alpha) = \nu(a_n, \alpha; \theta).$$

To ensure that all relevant type thresholds in the results below are in the support, I assume that θ_{max} is sufficiently large. Agents share action space $A_n = 0, 1, \ldots, M$ and the aggregator is mean action $\alpha(a) = \int_{n \in N} a_n d\mu(n)$. I focus on anonymous, unilateral, on-path budget-balanced linear taxes: $T_n(a_n) = -pa_n + p\alpha(a^*(p))$, where a policy is indexed by the tax rate p.

I impose the following monotonicity and single-crossing conditions:

Assumption 2 (Monotonicity)**.** *The utility function* ν *satisfies the following monotonicity properties:*

1. *For any θ and α, $\nu(a_n, \alpha; \theta)$ is strictly increasing in a_n.*

2. *For any θ and a_n, $\nu(a_n, \alpha; \theta)$ is weakly decreasing in α, and there exists $h < M$ such that if $a_n > h$ it is strictly decreasing.*

Assumption 2 essentially implies that privately beneficial actions impose social harm, which generates a wedge between the unregulated equilibrium and the social optimum. This creates scope for taxation to improve welfare.

Assumption 3 (Strong Single Crossing)**.** *The utility function ν has strictly bounded increasing differences in (θ, a_n) and (θ, α); i.e., there exists $\epsilon > 0$ such that:*

1. *For any $a_n > a_n'$, α, and $\theta > \theta'$,*

$$[\nu(a_n, \alpha; \theta) - \nu(a_n', \alpha; \theta)] - [\nu(a_n, \alpha; \theta') - \nu(a_n', \alpha; \theta')] \geq \epsilon(\theta - \theta').$$

2. *For any $\alpha < \alpha'$, $a_n > h$, and $\theta > \theta'$,*

$$[\nu(a_n, \alpha; \theta) - \nu(a_n, \alpha'; \theta)] - [\nu(a_n, \alpha; \theta') - \nu(a_n, \alpha'; \theta')] \geq \epsilon(\theta - \theta').$$

Assumption 3 is a standard single-crossing/increasing-differences condition, strengthened to bound the increase away from zero, ensuring that preferences are sufficiently responsive to types. Its first part ensures that for any tax t, high actions are chosen by some sufficiently high types, and its second part ensures that a first-order increase in type translates to a first-order increase in the value of reducing the externality. Models with preferences linear in type—where $\nu(a_n, \alpha; \theta) = \theta \cdot \hat{\nu}(a_n, \alpha)$ for some type-independent $\hat{w}$—satisfy this requirement by construction.

The next result shows that, under these assumptions, any budget-balanced linear tax that improves on the status quo generates spurious opposition:

Proposition 3 (Spurious opposition)**.** *Let $\mathcal{E}$ be an aggregative economy. Suppose Assumptions 2–3 hold and fix a status-quo policy $T^0 = p^0$ with associated equilibrium aggregate α^0. Any budget-balanced linear tax p that induces an equilibrium aggregate $\alpha^p < \alpha^0$ generates spurious opposition: there exist θ^{EN} and $\bar{\eta}(\theta) \in (0, 1)$ such that for all agents with $\theta_n \in [\theta^{EN}, \theta_{max}]$ and all $\eta_n \in [0, \bar{\eta}(\theta)]$,*

$$\delta_n(T, T^0) > 0 \quad \textit{but} \quad \tilde{\delta}_n(T, T^0; \eta_n) < 0.$$

It is instructive to express this result in terms of its implications for weighted-support-based feasibility rules:

Corollary 1. *Let f be a weighted-support rule with threshold β, and assume $\eta_n = \eta$. Under the assumptions of Proposition 3, there exist a nonempty set $\mathcal{B} \subset (0, 1]$ and a threshold $\overline{\eta} < 1$ such that for all $\beta \in \mathcal{B}$:*

$$f(\tilde{\delta}(T, T^0; \boldsymbol{\eta})) = 1 \iff \eta > \bar{\eta}$$

In words, for some feasibility thresholds, the socially beneficial taxation p is feasible only if neglect is not too large.

The intuition for Proposition 3 is simple: budget balanced taxation always creates net payers. When preferences are sufficiently rich, if switching costs—i.e., the private value of taking the socially harmful action—and externality exposure—i.e., the value of reducing α—lie on the same dimension (or equivalently, if there is a monotonic mapping from costs to externality exposure; this is guaranteed by Assumption 3), then, despite being net payers, high types' actual gains remain positive. Provided their EN parameter η is sufficiently low, this generates spurious opposition.

Proposition 3 delivers three main takeaways on the prevalence of EN, on the types it affects, and on the economic conditions for it to constrain feasibility. First, it shows that the constraint EN imposes on feasibility extends beyond the contrived examples discussed earlier in this section, and arises naturally in common economic settings.

Second, the proposition identifies high types—typically interpreted as those with higher skill or wealth—as the set of agents most likely to spuriously oppose a policy. To the extent that skill and wealth correlate with greater influence weight, this makes spurious opposition more likely to affect feasibility.[13]

Last, the proposition provides an intuitive sufficient condition on the economic settings where EN is more likely to constrain feasibility: spurious opposition emerges when high value of taking the socially harmful action coincide with large equilibrium benefits from reducing the externality. Put differently, settings particularly prone to spurious oppositions are those where high usage is associated with greater exposure to equilibrium effects.

For example, EN is unlikely to pose a feasibility challenge when a few agents (e.g., factory owners) value a harmful activity (e.g., polluting a river), but the benefit from its reduction is concentrated among a different group (e.g., downstream residents). Here, while taxing pollution may generate opposition from factory owners, it need not be spurious, but driven by real economic interest.

Examples where EN is likely to constrain feasibility include dynamic electricity pricing (Joskow and Wolfram 2012), where high-usage consumers face the steepest rate increases but gain most from grid stability; government-provided deposit insurance financed by

13. More speculatively, higher types, whose behavior is less responsive to the tax, may project their behavior onto others and exhibit lower η, which may exacerbate the constraint still further.

a risk-based fee (Ronn and Verma 1986), where banks with the riskiest portfolios pay the highest premiums but, as the institutions most vulnerable to bank runs (Amador and Bianchi 2024), also gain the most from preventing them; and, finally, as my running example illustrates, in the context of congestion pricing, where those who place the highest value on driving are also those who place the highest value on the shorter commute times that result from the toll.

Example (Politically Infeasible Congestion Pricing.) From this point on, I fix I and suppress it from the notation. First of all, notice that (fixing I) the congestion model is an aggregative economy, where agents' utility depend on their own choices, an the aggregate level of congestion, and a tax/transfer. Second, since value of time is linear in type, the example satisfies Assumptions 2 (with $h = 0$) and 3 by construction.[14]

As is common in settings with externality, without a toll ($p = 0$), congestion is inefficiently high: the threshold for driving in the socially optimal allocation is $\theta^* > \theta^0$.

Result 2. *In the congestion model, without tolls ($p = 0$), the equilibrium driving rate exceeds the social optimum: $\alpha^0 > \alpha^*$. Moreover, there exists a toll $p^* > 0$ that implements the social optimum.*

The canonical solution for this inefficiency is congestion pricing (Pigou 1920; Vickrey 1969). A budget-balanced congestion pricing is a toll that implements the socially optimum allocation, coupled with a lump-sum rebate k redistributing the revenue. In the proof of Result 2, I derive the optimal toll:

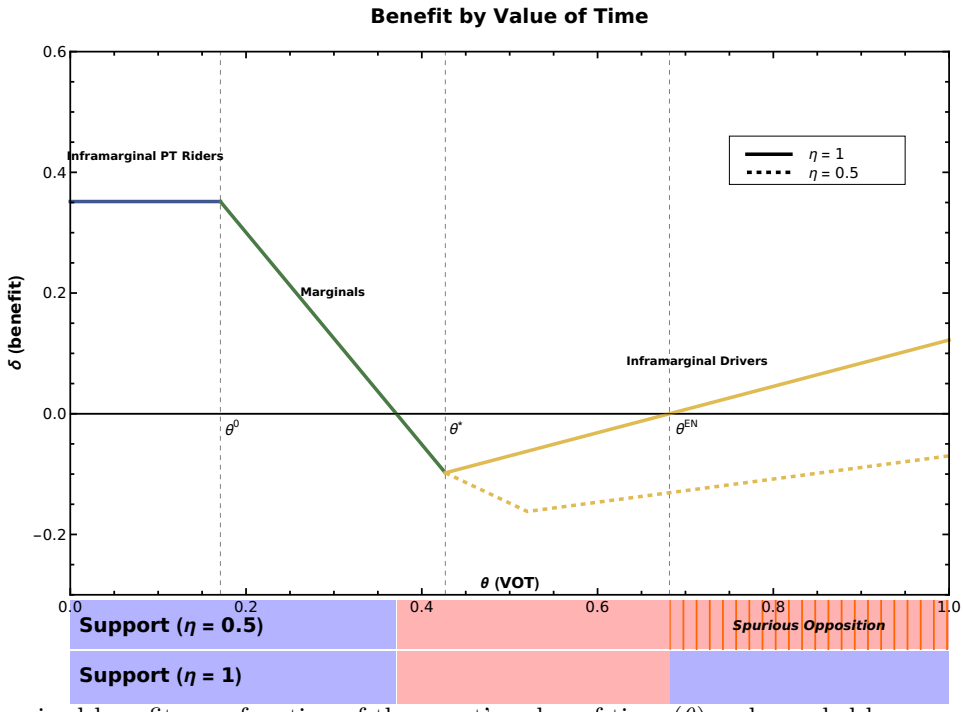
$$p^* = -Q'(1 - F(\theta^*)) \cdot \mathbb{E}[\theta | \theta \geq \theta^*] \cdot (1 - F(\theta^*)),$$

where θ^* is the driving cutoff at the optimum. As is typical in these settings, the optimal toll is the externality: the effect of adding an additional driver factored by the mean value of time among drivers and the size of the driving population. The budget-balanced transfer is therefore $\alpha^* p^*$. Let $Q(\alpha^0) = Q^0$ and $Q(\alpha^*) = Q^*$.

Who are the actual winners and losers from congestion pricing? The solid line in Figure 1 plots gains from congestion pricing as a function of type. Three classes of commuters emerge: the leftmost flat segment in blue contains inframarginal public-transit riders ($\theta \leq \theta^0$); agents in this group pay no toll and do not adjust their behavior, and their gain is simply the flat transfer $\alpha^* p^*$. The middle, downward-sloping segment in green contains the marginals ($\theta \in (\theta^0, \theta^*]$)—those who switch to public transit in response to the toll. The gain for these commuters is $\alpha^* p^* - \theta Q^*$—the size of the transfer minus the cost of switching to public transit, which increases with their value of time. Finally, the rightmost, upward-sloping segment in yellow represents inframarginal drivers

14. Strictly speaking, for the example to be a special case of the setting of Proposition 3, b would need to be a part of the status-quo policy (i.e., $p^0 = b$.)

Figure 1: Congestion Pricing: Perceived vs. Actual Benefit



Note: Perceived benefit as a function of the agent's value of time (θ), color coded by group (marginal vs. inframarginal). The solid line represent perfect foresight ($\eta = 1$), and the dashed line represents EN ($\eta = 0.5$). The vertical dashed lines mark the cutoff values θ_0 (between inframarginal PT riders and marginals) and θ^* (between marginals and inframarginal drivers). Solid bars at the bottom indicate positive (blue) and negative (red) perceived benefit (i.e., support/opposition) under perfect foresight (bottom row) and EN (top row), and the striped segment ("Spurious Opposition") indicates where the two diverge. The status quo has no toll and no transfer. The proposed policy is a budget-balanced congestion pricing with a toll p^* implementing the social optimum (α^*) and rebates equilibrium toll revenue ($\alpha^* p^*$) as a flat transfer. Parameters: $Q(\alpha) = 1.5\alpha$, $b = 0.3$ and $\theta_n \sim U(0, 1)$.

who pay the toll and continue driving ($\theta > \theta^*$). For this group, the actual gain is $-p^*(1 - \alpha^*) + \theta(Q^* - Q^0)$: the flat transfer minus the toll—which is negative under a budget-balanced policy—plus the increase in time savings, scaled by their value of time.

The figure shows that real gains from congestion pricing are typically non-monotonic in θ: for low-θ agents (all inframarginal PT riders and lower-θ marginals) driving has low enough value to be fully offset by the transfer. Medium-θ agents lose: for marginals, driving is valuable enough that the transfer cannot offset the cost of switching, while for lower-θ inframarginal drivers, the value of time savings is too small to cover the toll. Finally, agents with high θ value time savings enough to come out ahead after paying the toll (net of the transfer).

The dashed curve in Figure 1 plots *perceived gains*, for $\eta = 0.5$. Inframarginal PT riders and marginals do not drive in equilibrium under congestion pricing, therefore their gains are independent of congestion, making perceived and actual gains coincide. For

inframarginal drivers, perceived gains are:

$$\tilde{\delta}_n(\tilde{T}, T^0; \eta_n) = \underbrace{p^*(\alpha^* - \hat{a}_n)}_{T^*(\hat{a}_n)} - \underbrace{\big(1 - \hat{a}_n\big)(\theta_n Q^0 - b)}_{w_n(\hat{a}_n, a^0_{-n}) - w_n(a^0)} + \eta_n \big[\underbrace{\theta_n(Q^* - \hat{a}_n Q^0)}_{w_n(a^*) - w_n(a^0)} - \underbrace{(1 - \hat{a}_n)(p^* + b)}_{T^*(a^*_n) - T^*(\hat{a}_n)} \big],$$

where $\hat{a}_n = \mathbf{1}_{\theta_n Q^0 \geq p^* + b}$ is the optimal action under the new toll and the old congestion level. The direct effect is the transfer minus the toll or the switching cost (depending on $\hat{a}_n$), and the equilibrium effect is the new time saving net the old time savings or the toll-inclusive driving cost (also depending on $\hat{a}_n$).

As the yellow dashed line in Figure 1 shows, the attenuation of the congestion-reduction effect initially makes inframarginal drivers perceive themselves as potential marginals ($\hat{a}_n = 0$), which causes perceived benefits to *decrease* with θ (as for marginals). Once θ is high enough that they expect to remain inframarginals even when congestion persists (i.e., $\hat{a}_n = 1$), perceived benefit increases again with value of time, but at a slower rate.

The spurious opposition includes all agents to the right of where the solid line meets the x-axis, denoted θ^{EN} (from Proposition 3)—agents with positive real gains, whose perceived gains are negative under sufficient neglect. For this group, perceived gains are positive iff

$$\eta \geq \frac{p^*(1 - \alpha^*)}{\theta_n(Q^* - Q^0)} \equiv \overline{\eta}(\theta_n),$$

i.e., when η is higher than the toll-to-equilibrium-benefit ratio. Consequently, when inframarginal drivers are pivotal, EN can make congestion pricing politically infeasible. For example, with equal weights, whenever $0 < S(\delta) - \beta < 1 - F(\theta^{EN})$, congestion pricing is feasible if and only if η is sufficiently high.

3.5 Remedy

With standard tools—anonymous, unilateral and budget-balanced taxation—the political feasibility of any outcome a social planner wishes to implement may be constrained by EN. Can alternative tools help overcome these constraints?

This section constructs a remedy that applies in generic quasi-linear economies. Our objective is to construct a politically feasible policy that implements a social outcome a under fiscal constraints. Formally, given a status quo T^0, a target action profile $a \in A$, a maximal deficit level $\overline{D}$, and a feasibility rule f, we are looking for a transfer scheme T that satisfies the following criteria:

1. *Outcome implementation:* $a^*(T) = a$.
2. *Budget constraint:* $D^*_T \leq \overline{D}$.
3. *Political feasibility:* $f(\tilde{\delta}(T, T^0; \boldsymbol{\eta})) = 1$.

Notice that achieving these goals may be impossible for reasons entirely unrelated to EN. Indeed, there is a substantial body of work on the challenges in achieving each criterion: mechanism design, public finance and social choice theory study implementation, fiscal constraints, and preference aggregation, presenting a rich set of challenges even absent any cognitive constraints. To isolate EN from these standard frictions, it is instructive to start with a policy that implements some social outcome a under the budget constraint, while being politically feasible under perfect foresight, i.e., $f(\delta(T, T^0)) = 1$.

3.5.1 EN-robust Policies

Consider a policy T with associated equilibrium a^T and deficit D. A spurious opposition arises when $v_n((\hat{a}_n, a^0_{-n}); T)$ (the agent's utility from T holding others' behavior fixed at a^0), is significantly lower than $v_n(a^T; T)$ (the agents actual utility under T). We therefore seek a modified policy $\tilde{T}$ that closes this gap while inducing the same equilibrium and satisfying the same budget constraint as T.

As it turns out, a multilateral policy $\tilde{T}$ that supplements T with a compensation k contingent on the profile a and pays this compensation whenever $a = a^0$ a.e. can achieve exactly these three goals. To see why, observe that this compensation depends only on others' actions, and therefore with quasi-linear utilities has no effect on each agent's choice of a_n. It must therefore result in the same equilibrium as T. This also implies that the two policies T and $\tilde{T}$ have the same deficit, as the compensation is never paid in equilibrium. Finally, since $v_n((\hat{a}_n, a^0_{-n}); \tilde{T}) = v_n((\hat{a}_n, a^0_{-n}); T) + k$, a sufficiently large k can close the gap between $v_n((\hat{a}_n, a^0_{-n}); T)$ and $v_n(a^T; T)$.

The theorem below formalizes this intuition. I call two policies that induce the same Nash equilibrium *outcome-equivalent*, and two policies that have the same equilibrium deficit *deficit-equivalent*. Finally, action profiles a and a' are *distinct* if they differ for a set of positive measure.

Theorem 1. *Let $\mathcal{E}$ be a quasi-linear economy, T^0 a status-quo policy, f a monotonic feasibility rule, and T^* an anonymous unilateral policy. Suppose that $a^*(T^*)$ and $a^*(T^0)$ are distinct. If T^* is politically feasible under T^0 with respect to f when $\boldsymbol{\eta} = \mathbf{1}$, then there exists a multilateral policy $\tilde{T}$ that is outcome and deficit equivalent to T^*, and is politically feasible with respect to f for all $\boldsymbol{\eta} \in [0,1]^N$.*

A policy is *EN-robust* if it is politically feasible with respect to f for all $\boldsymbol{\eta} \in [0,1]^N$. When discussing policies that implement specific outcomes, "EN-robust" refers specifically to those that also satisfy the outcome- and deficit-equivalence requirements.

Theorem 1 shows that if a policy is politically feasible without equilibrium neglect, it has an EN-robust counterpart that maintains feasibility even in the presence of equilibrium-neglecting voters. The result imposes no structure on heterogeneity, requires no knowledge about the player other than what *others* did in the status quo, and applies quite

generally: given any monotonic feasibility rule, starting from any status quo, and for any constrained optimum achieved via transfers.[15] Crucially, the modified policy $\tilde{T}$ generates the same equilibrium as the original policy T^*, without increasing the deficit on the equilibrium path.

3.5.2 Off-path Deficit

While contingent compensations maintain the same deficit in equilibrium, they can increase it off the equilibrium path. A potential worry is that policies promising large contingent compensations may be met with skepticism by voters—especially equilibrium-neglecting voters, who believe these compensations will actually be triggered. Since the goal is precisely to persuade such neglectful voters, this skepticism may counteract the persuasive effect of the compensation. Moreover, policymakers, who may themselves suffer from equilibrium neglect, may reject policies with large off-path compensations to avoid potential large deficits. This raises a natural question: can contingent compensations be constructed such that the deficit is constrained not just in equilibrium, but at every action profile?

The theorem below shows that any policy T has an EN-robust counterpart $\tilde{T}$ that maintains off-path budget constraint in the following sense: at the status-quo profile, the deficit equals the status-quo deficit, and at any other action profile, it equals the deficit of T. However, constructing such policies requires additional information: the size of the compensation must be conditioned on each player's status-quo action.

Formally, let T^0 be a status-quo policy with associated equilibrium $a^0 = a^*(T^0)$. A policy T is past-action based (PA-based) if transfers to each player depend not only on current actions but also on that player's action under the status quo: $T_n(a) = T'_n(a; a^0_n)$ where a^0_n is player n's action under T^0 treated as a fixed parameter.

Theorem 2. *Let $\mathcal{E}$ be a quasi-linear economy, T^0 a status-quo policy, f a weighted-support-based rule, and T^* an anonymous unilateral policy. Suppose that $a^*(T^*)$ and $a^*(T^0)$ are distinct. If T^* is politically feasible under T^0 with respect to f when $\boldsymbol{\eta} = \mathbf{1}$, then there exists a PA-based multilateral policy $\tilde{T}$ that is outcome and deficit equivalent to T^*, and is politically feasible with respect to f for all $\boldsymbol{\eta} \in [0,1]^N$. Furthermore, $\tilde{T}$ may be constructed so that its deficit equals that of the status quo at a^0 and that of T^* otherwise: $D_{\tilde{T}}(a^0) = D_{T^0}(a^0)$ and $D_{\tilde{T}}(a) = D_{T^*}(a)$ for all $a \in A \setminus a^0$.*

Notice that unlike Theorem 1, EN-robustness here is guaranteed only with respect to weighted-support-based rules, a distinction I discuss below.

15. While theorems 1 and 2 are stated for economies with a continuum of agents, their proofs show how the results extend to finite populations under the appropriate notion of distinctness, requiring action profiles to differ for more than one player.

The contingent compensation that maintains the status-quo deficit at the status-quo profile depends on past actions, and is given by

$$T_n^0(a^0) - T_n^*(a_n^0), \tag{2}$$

the difference between what agent n pays under the status quo and what they would pay under the new policy if they maintained their status-quo behavior. Looking at (2), it is straightforward to see that summing the transfers over all agents, we get

$$D_{\tilde{T}}(a^0) = D_{T^*}(a^0) + D_{T^0}(a^0) - D_{T^*}(a^0) = D_{T^0}^*.$$

Crucially, the expression in (2) only requires knowledge of the agent's past actions.[16] While seemingly challenging in practice, past actions can often be well proxied by readily available observables. For example, in my empirical application I show that workplace location can be used to predict driving frequency.

The intuition for Theorem 2 combines two main insights. First, when the population is non-responsive to taxation, revenue is higher, funding the off-path compensation. Second, knowing an agent's status-quo action provides an upper bound on their direct harm from the proposed policy, enabling more precise targeting of potential members of the spurious opposition. In congestion pricing, for example, high congestion generates more toll revenue, and an agent who drives in the status quo is harmed by no more than the toll size (an agent taking public transit is not harmed at all). The policy thus adds a contingent compensation that targets those directly harmed, precisely offsetting the direct effect. When compensations are triggered they are funded by the revenue surplus, leading to the same deficit as the status-quo deficit. As an immediate implication, when a policy T^* is not intended to raise revenue (i.e., $D(T^*(a^*)) \leq D(T^0(a^0))$), its PA-based, EN-robust counterpart $\tilde{T}$ maintains the same deficit as T^* in all contingencies.

One implication of past-action based targeting is that when $\eta < 1$, some agents perceive lower benefits under $\tilde{T}$ than under T^*. To see why, observe that the contingent compensation in $\tilde{T}$, given by (2), can be negative. Therefore, while the contingent compensation ensures that all agents with positive actual gains from T^* support $\tilde{T}$ over the status quo, it may lead some of them to support $\tilde{T}$ less intensively than they would support T^*. This is precisely why Theorem 2 holds only for weighted-support-based feasibility rules, which depend on the sign of $\tilde{\delta}_n$, but not on its intensity. Consequently, the theory predicts that relative to its standard counterpart, the increase in the support for PA-based, EN-robust policy should be concentrated among potential recipients, i.e., those whose past actions imply higher contingent compensation.

16. In fact, expression (2) shows that the compensation depends only on what the agent paid previously and what they would pay under the new policy with unchanged behavior, which may be a strictly weaker informational requirement than knowing the agent's status-quo action.

Aggregate Uncertainty Appendix B extends Theorems 1 and 2 to aggregative economies with uncertainty, where a policy induces a distribution over equilibrium outcomes rather than a single outcome. It shows that every policy has an EN-robust counterpart that induces the same *distribution* over equilibrium aggregates while maintaining the same *expected* deficit. To achieve this, the contingent compensations scale linearly with the distance between the realized and expected equilibrium aggregate, creating a lottery with positive perceived expected value for equilibrium-neglecting agents while maintaining the same expected deficit under the true distribution, thus extending Theorem 1. When compensations can be conditioned on an agent's *distribution* of past actions (rather than a single deterministic action), we can further ensure that, under the status-quo distribution, the compensations have the same expected deficit as the status-quo deficit, thus extending Theorem 2.

Example (Congestion-contingent compensation.) To see the difference between Theorems 1 and 2 more clearly, we can apply them in the congestion model.

Pseudo Rebates Consider the anonymous multilateral congestion pricing policy $\tilde{T}$ that augments T^* with contingent compensation $\tilde{k}(a)$ that pays when congestion remains high:

$$\tilde{T}(a) = \underbrace{p^*(\alpha^* - a_n)}_{T^*(a_n)} + \underbrace{\mathbf{1}\{\alpha(a) = \alpha^0\}p^*}_{\tilde{k}(a)},$$

In addition to the toll p^*, and the flat transfer $k^* = \alpha^* p^*$ (both identical to T^*), this policy compensates all agents p^* when congestion remains high ($\alpha(a) = \alpha^0$), effectively rebating the toll. This rebate completely eliminates the spurious opposition. To see why, recall that only inframarginal drivers are in the spurious opposition. For these drivers:

$$\begin{aligned}\tilde{\delta}_n(\tilde{T}, T^0; \eta_n) = \\ &\underbrace{p^*(\alpha^* - \hat{a}_n)}_{T^*(\hat{a}_n)} - \underbrace{(1-\hat{a}_n)(\theta_n Q^0 - b)}_{w_n(\hat{a}_n, a^0_{-n}) - w_n(a^0)} + \underbrace{p^*}_{\tilde{k}(a^0)} + \eta_n \big[\underbrace{\theta_n(Q^* - \hat{a}_n Q^0)}_{w_n(a^*) - w_n(a^0)} - \underbrace{(1-\hat{a}_n)(p^* + b)}_{T^*(a^*_n) - T^*(\hat{a}_n)} - \underbrace{p^*}_{\tilde{k}(a^0)} \big] \\ &= \eta \left[\theta_n(Q^* - Q^0) - (1-\alpha^*)p^*\right] + (1-\eta)\left[p^*\alpha^* + (1-\hat{a}_n)(p^* + b - \theta_n Q^0)\right].\end{aligned}$$

The first term is positive for all drivers with positive real gains, while the second term is always positive (since $\hat{a}_n = 0 \iff p^* + b \geq \theta_n Q^0$). Intuitively, the direct harm is bounded by the size of the toll, and the rebate completely offsets it. Notice that the rebate only depends on the aggregate level of congestion, making it a *pseudo rebate* since it is independent of the commuter's own driving choice. Therefore, the effective toll under $\tilde{T}$ remains p^*, the equilibrium level of congestion remains α^* and the compensation

is not triggered in equilibrium. Since T^* is budget-balanced in equilibrium, so is $\tilde{T}$: $D^*_{\tilde{T}} = D^*_{T^*} = 0$. This illustrates Theorem 1.

However, at a^0, the deficit is

$$D_{\tilde{T}}(a^0) = \underbrace{(1+\alpha^*)p^*}_{\text{transfers}} - \underbrace{\alpha^0 p^*}_{\text{revenue}} > 0,$$

i.e., off equilibrium, the policy runs a deficit.[17]

Commuter Compensation Now consider the PA-based multilateral policy $\tilde{T}$:

$$\tilde{T}_n(a) = \underbrace{p^*(\alpha^* - a_n)}_{T^*(a_n)} + \mathbf{1}\{\alpha(a) = \alpha^0\} \underbrace{p^*(a_n^0 - \alpha^*)}_{T^0(a_n^0) - T^*(a_n^0)},$$

This *commuter compensation* policy compensates status-quo drivers $p^*(1-\alpha^*)$ when congestion remains high despite having a toll in place. Following the same derivation, for inframarginal drivers (all status-quo drivers who receive compensation), if $\delta_n(T^*, T^0) \geq 0$, then $\tilde{\delta}_n(\tilde{T}, T^0; \eta) \geq 0$ for all $\eta \in [0,1]$.[18] The same outcome- and deficit-equivalence arguments apply. Since α^0 is the share of commuters with $a_n^0 = 1$, the deficit at a^0 is:

$$D_{\tilde{T}}(a^0) = \underbrace{\underbrace{\alpha^* p^*}_{k^*} + \underbrace{(1-\alpha^0)}_{\text{share } a_n^0=0}(0-\alpha^*)p^* + \underbrace{\alpha^0}_{\text{share } a_n^0=1}(1-\alpha^*)p^*}_{\text{transfers}} - \underbrace{\alpha^0 p^*}_{\text{revenue}} = 0.$$

Therefore, a contingent compensation targeting status-quo drivers can be constructed such that at a^0, aggregate transfers equal the revenue. This illustrates Theorem 2.

3.5.3 Discussion

Transfers can help increase political support. The challenge is in designing a transfer that is both credible to agents and costless to the planner. A large payment contingent on an event all parties deem unlikely (e.g., snow in the Sahara) has little persuasive power, while one contingent on a likely outcome (bad weather day in Boston) may be prohibitively expensive. The mechanism I propose operates precisely on the wedge generated by equilibrium neglect: it identifies the contingency—the persistence of status-quo behavior—that EN agents find plausible (and/or salient), but the planner knows lies off the new equilibrium path. Much like a warranty, the planner (as producer) attaches to the policy (as product) a contingent guarantee (warranty/insurance) that pays if it fails

17. Setting $\tilde{k}(a^0) = (1-\alpha^*)p^*$, i.e., making the *total* transfer at a^0 equal the toll p^*, still yields a deficit $p^*(1-\alpha^0) > 0$ which is positive whenever inframarginal PT riders exist.

18. As discussed above, inframarginal PT may prefer the original policy. For these agents, $\tilde{\delta}_n(\tilde{T}, T^0; \eta) \leq \tilde{\delta}_n(T^*, T^0; \eta)$ with strict inequality for $\eta < 1$. However, $\delta_n(T^*, T^0) \geq 0 \Rightarrow \tilde{\delta}_n(\tilde{T}, T^0; \eta) \geq 0$ still holds, ensuring EN-robustness under any weighted-support-based feasibility rule.

to deliver the promised outcome. And just as a producer can offer the warranty because of their knowledge of the true product quality, the planner can offer the compensation because they know the actual equilibrium.[19] Notice, however, that unlike a warranty, compensation here depends on aggregate outcomes that no single agent can influence, so it creates no moral hazard.

It is instructive to compare this approach to more conventional remedies for political opposition, viewed through the lens of congestion pricing. Decreasing the toll (as was recently done in New York City) may increase support, but would likely violate both outcome equivalence (by inducing a higher level of driving compared to the social optimum) and generate lower revenue. A planner could use subsidies instead of tolls to achieve the same goal (e.g., by paying drivers to drive outside of peak hours). However, this would violate the budget constraint. An alternative policy, seemingly similar to contingent compensation, is refunding drivers the tolls they paid if congestion remains high. However, this approach effectively rewards driving, which distorts incentives and violates outcome equivalence. Finally, trials and pilot programs, such as the one used to introduce congestion pricing in Stockholm (Schuitema et al. 2010), offer another alternative by making equilibrium benefits tangible and directly observable, allowing direct experience to correct the perceptions that fuel spurious opposition (the next section presents global evidence consistent with such corrections). However, ultimately, this approach faces the same fundamental feasibility challenge: if a policy is politically infeasible to implement permanently, it may also be infeasible to trial, as agents who expect to lose from it have no incentive to agree to a temporary version. The off-path compensation mechanism avoids this trap by offering compensations that undo the payments when the equilibrium prediction does not pan out.

A potential worry is that state-contingent policies are themselves too complex. If agents neglect equilibrium effects due to cognitive constraints, they may also struggle to process the contingent nature of the proposed remedy. This raises an important empirical question: can EN-robust policies increase support in practice? I address this question in Section 4.

4 Congestion Pricing: Public Opinion

In this section I test two of the main components of my model in the context of congestion pricing: first, that voters under-appreciate its equilibrium effects; second, that contingent compensation that pays when congestion remains high increases public support for the

19. Warranties can operate either directly as insurance that eliminates risk, or as signals that affect beliefs about product quality. In the next section I empirically test the effect of contingent compensation on beliefs. For evidence in consumer settings, see Boulding and Kirmani (1993) and Chu and Chintagunta (2011).

policy.

As discussed in the introduction, congestion pricing is a natural settings for studying equilibrium neglect and testing EN-robustness, as it offers a clear distinction between the direct effect (toll increase) and equilibrium effect (congestion reduction). In addition aggregate outcomes (e.g. traffic volume) can be easily and transparently measured, making contingent payments simpler and more credible. Last, the relative simplicity of the setting makes predictions quite reliable.[20] For example, in its final Environmental Assessment document, the Metropolitan Transportation Authority (MTA) predicted a 13% decrease in daily entrances to the Central Business in Manhattan under the $9 tolling scenario. As of September 7th, the average weekly decrease is 12%.[21] I start by reviewing motivating evidence from past polls.

4.1 Motivating Evidence from Existing Polls

As motivation, I start by analyzing existing polls. Figure 2 plots *net support*—the percentage supporting minus the percentage opposing the policy—before versus after implementation of congestion pricing in seven different cities (to the best of my knowledge, these are all the cities for such data exist). Support consistently increases post-implementation. To the extent that pre-implementation responses reflect perceived gains and post-implementation responses reflect respondents' actual experience (i.e., actual gains), these results are consistent with equilibrium neglect generating spurious opposition that dissipates post-implementation, once benefits realize.

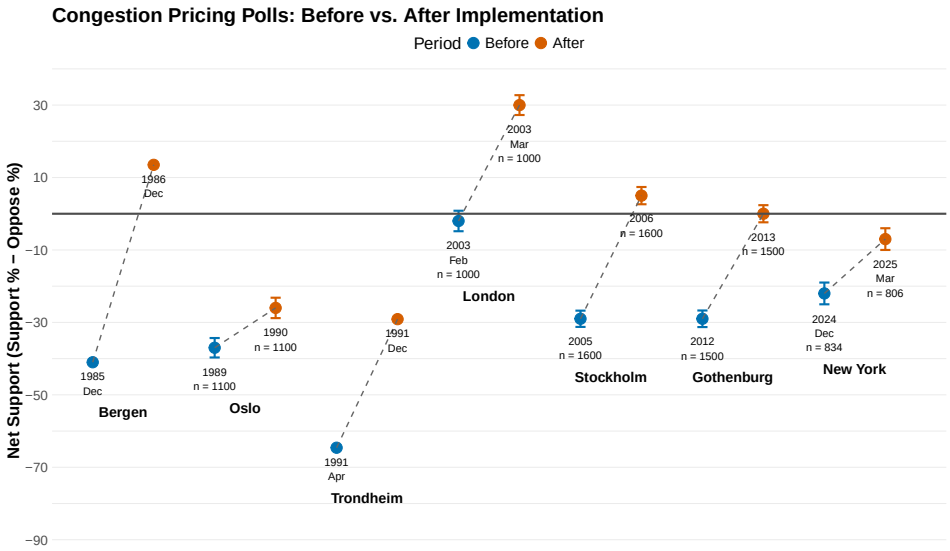
To show this channel more directly, I analyze individual responses from New York ($N = 1,640$), which I obtained from Siena Polls. I focus on two polls: one conducted one month before congestion pricing began in NYC (December 2–5, 2024; $N = 834$) and another about two months after (March 2–6, 2025; $N = 806$), both sampling registered voters from New York state using identical polling methodology. For each respondent, I observe their support response and a rich set of demographics, including county of residence. If attitude shifts reflect belief updating induced by policy exposure, respondents less likely to commute to Manhattan during the two months between implementation and the second poll should also be less likely to update their views. Using ACS county-level commuting flow data, I calculate a Manhattan commute intensity measure for each county, defined as the share of that county's commuters who commute to Manhattan (NY county) out of all commuters in the county. I classify respondents as high-commute (above-median commute intensity) or low-commute (below-median).

Figure 3 plots net support before and after implementation separately for each com-

20. Theoretically, congestion games have a potential function which guarantees convergence to a Nash equilibrium in pure strategies via best-response dynamics.

21. For the MTA prediction, see: https://www.mta.info/document/158191. For current data on entries, see: https://metrics.mta.info/?cbdtp/vehiclereductions

Figure 2: Support for Congestion Pricing Across Cities



Note: Support for congestion pricing pre- (blue) and post- (orange) implementation across cities. Y-axis: net support (% support - % oppose). Aggregate data assembled from various publicly available sources, listed in Online Appendix Table C.1. Error bars represent standard errors, and are added when the number of observations was provided.

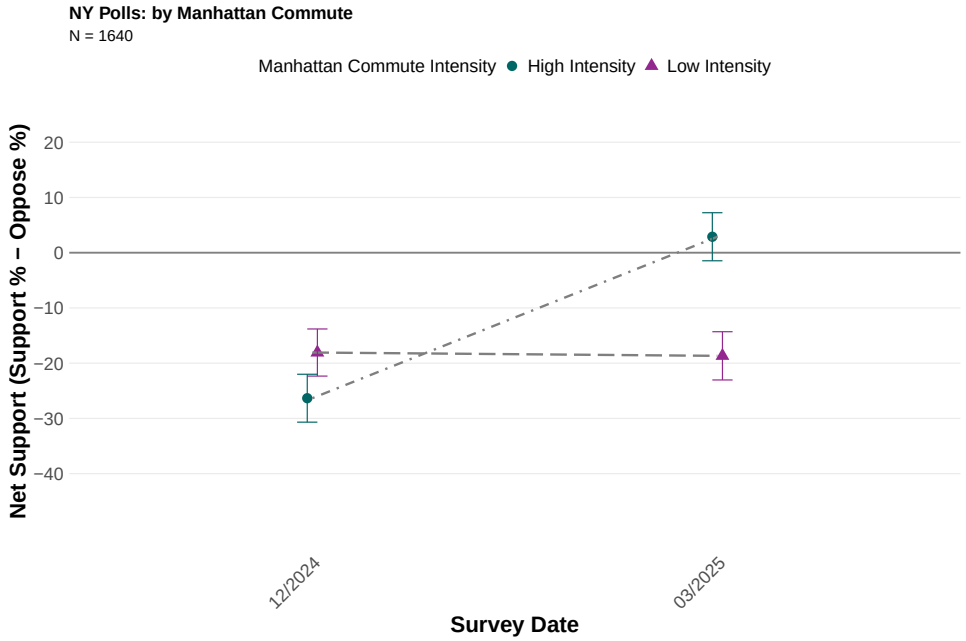
mute intensity group. It shows that the increase in support in NYC is driven by high-commute respondents. This group started with slightly lower level of support, but increased by 29.8 percentage points more (SE = 8.48) after implementation. Appendix Tables C.3 and C.4 show that the difference remains large and significant without sampling weights, controlling for demographics, when fully interacting demographics with an indicator for post-implementation,[22] and when including three additional polls by Siena, one conducted before the policy took effect, and two after (see table C.2). While question wording differs across waves,[23] the differential increase suggests policy exposure plays an important role in driving attitude shifts, consistent with respondents holding overly pessimistic beliefs that change once they experience the gains from congestion pricing. In the next section, I provide more direct evidence that over-pessimism shapes opposition

22. Differential partisan dynamics are a particularly worrying confound since the policy became more salient after President Trump took office in January 2025. Indeed, the fully saturated model in Tables C.3 and C.4 shows that Republicans became more negative towards congestion pricing after its implementation, but the differential increase among high-commute participants becomes only slightly smaller and remains highly statistically significant.

23. See Table C.2. The pre-implementation question reads: "In June, Governor Hochul put a $15 congestion pricing toll plan for Manhattan on hold. Last month, after the election, Governor Hochul reinstituted a $9 congestion pricing plan, set to begin in January. Do you support or oppose the Governor's decision to implement the $9 congestion pricing toll for Manhattan?" with the options Support/Oppose/In the middle, whereas the post-implementation question reads: "Governor Hochul says that congestion pricing in Manhattan is working, reducing traffic and raising revenue for the MTA, and it should remain. President Trump says that congestion pricing is nothing more than an unfair tax on working commuters, and it should be eliminated. Do you think congestion pricing:" with the options should be eliminated/should remain/in the middle.

using my own survey.

Figure 3: Support for Congestion Pricing in NY by Manhattan Commute Intensity



Note: Support for congestion pricing in NYC among NY registered voters pre- and post-implementation, split by commute intensity to Manhattan. Total $N = 1,640$. Sampling weights applied. High intensity (green circles): respondents residing in counties above median Manhattan commute share. Low intensity (purple triangles): below median. Commute intensity: commuters traveling to Manhattan (NY county) as share of all commuters, based on 2020 ACS commuting flows data. Error bars represent standard errors. Diff-in-diff estimate: 29.83 (SE = 8.48).

4.2 Survey Design

The survey has two main purposes, testing two main components of the theory. The first is descriptive, testing for the presence of equilibrium neglect (in the form of congestion-reduction skepticism) in an ecologically important high-stakes policy setting, and then showing how it maps to policy stances in a representative sample. The second is testing whether the remedy suggested by the theory, contingent compensation, is indeed an effective tool for increasing public support.

The survey follows a simple two-stage, between-subject design: subjects first indicate support for a standard congestion pricing policy, answer beliefs and habits questions, and then in the second stage indicate support for a randomly assigned modified congestion pricing policy. These two stages correspond to the two goals stated above: directly documenting equilibrium neglect, and then testing the effectiveness of EN robustness. The survey instruments are available in Appendix D.

4.2.1 Baseline support

At the beginning of the survey, all respondents are presented with a standard congestion pricing policy in the major downtown area of their metro area. According to the policy, modeled after the originally proposed NYC Central Business District Tolling Program, vehicles entering the downtown area during peak hours would be charged a $15 daily toll, and the revenue would be used for transportation improvements.

Respondents are presented with the arguments made by proponents (citing research from other cities suggesting it can cut traffic by about 15–20% and reduce travel times by 10–15 minutes) and opponents (saying that the policy is an unfair burden that will not meaningfully reduce congestion). Adding supporters' arguments has two main purposes. First, it is meant to eliminate trivial information or attention frictions which can presumably be corrected by simple informational treatments (rather than the involved policy change I am proposing). Second, it provides respondents with a natural traffic-reduction benchmark which is exploited in the EN-robust condition in the second stage. Opponents' arguments are presented to avoid one-sided presentation.[24] Respondents indicate their level of support for the policy on a five-point scale: Strongly Support, Somewhat Support, Neither Support nor Oppose, Somewhat Oppose, and Strongly Oppose.

4.2.2 Commute Habits & Beliefs

After indicating support for the policy, respondents answer questions in two order-randomized blocks. Block 1 contains two questions about commute habits: (i) the location of their workplace (downtown, other part of town, suburbs, rural, and home), and (ii) how frequently they drive into the downtown area (from "Never" to "5 or more times per week"). The first question is meant to identify downtown workers, the recipient population of the contingent compensation in the EN-robust condition. The second question is used in constructing the perceived benefit index described below.

Block 2 elicits beliefs about the travel time reduction as a result of congestion pricing and the level of certainty in the prediction. The question asked in this block is randomized: half of the respondents are asked to predict how congestion pricing would affect their own travel time ("If congestion pricing were implemented according to the proposal you just saw, how do you think it would affect driving time from your home to downtown [city]?"). Respondents select from nine options organized in 5-minute increments, ranging from "Driving time would decrease by more than 15 minutes" to "Driving time would not change" to 'Driving time would *increase* by more than 15 minutes." The other half of the sample is asked to predict the effect congestion pricing had in NYC. This question asks

24. A balanced presentation in policy questions is standard practice (American Association for Public Opinion Research 2023), meant to prevent framing/experimenter demand effects which have been shown to persist under one-sided presentation of arguments (Schuman and Presser 1996).

respondents to estimate post-implementation travel times through the Holland Tunnel during rush hour. Respondents are told the pre-toll average commute was 23 minutes (compared to 8 minutes with no traffic), and select from seven 5-minute increments ranging from "10 minutes or less" to "36 minutes or more." All respondents then report their level of certainty (0% to 100%) in their prediction following the same format as Enke and Graeber (2023).

4.2.3 Second-Stage Support

After answering these questions, all respondents see a modified congestion pricing policy and are asked to state their support on the same scale. The policy they see is randomly assigned. It is identical to the policy presented at baseline, with one of the following modifications, where the modification is highlighted for the respondent.

EN-robust (40% of the sample) The policy adds a "Monitoring & Assurance" clause: "Traffic levels would be monitored and published monthly. If monthly traffic decreases by less than 8% (about half the typical 15–20% reduction) all workers who commute to work in the downtown area would receive $120 tax credit for that month, regardless of how they commute." Since understanding the compensation mechanism relies on contingent reasoning—which equilibrium-neglecting respondents may struggle with—the treatment adds an explanation meant to explain the logic and reduce confusion: "This is intended to ensure affected workers either benefit from shorter commute times or receive compensation if traffic does not decline as expected."

Control (40% of the sample) The policy only includes the "monitoring" part without the contingent compensation ("Traffic levels would be monitored and published monthly"). To make the policy descriptions comparable, this treatment also included a similar explanation: "This is intended to provide data about the program's effects on traffic patterns."

Toll Reduction (20% of the sample): The toll is reduced from $15 to $9.

Discussion Theorem 2 from the theory section identifies status-quo commuters, those who currently drive into the tolled area, as the recipients of the contingent compensation in the budget balanced EN-robust policies. Since conditioning on past behavior may be impractical and may seem unnatural, the contingent compensation in the EN-robust condition targets downtown workers, an approximation meant to identify those who would be most directly affected by the policy. While not all downtown workers drive, and not all frequent drivers work downtown, downtown workers include many regular commuters

into the downtown area and represents a natural, clearly defined, administratively feasible group for implementing the compensation.[25]

To avoid knife-edge cases, rather than conditioning the compensation on zero improvement in traffic conditions, I condition it on half the predicted reduction. One may think of this version as an approximate budget-balanced version of the exact budget-balanced version proposed in the theory section.

Comparing 2nd-stage support between treatment and control conditions (controlling for baseline support) isolates the effect of contingent compensation itself, since the control condition absorbs any potential confounding factors including demand effects, repeated elicitation effects, and the effect of increasing perceived information/transparency from monitoring. This comparison therefore provides a lower bound on the real-world effect of introducing compensation into an existing congestion pricing policy, since doing so would also require introducing the monitoring itself.

Finally, the toll-reduction condition is meant to serve as a benchmark. In the face of strong opposition, a natural reaction by a politician is to lower the toll to appease the opposition. This condition is therefore included to test off-path compensation against a more conventional adjustment margin. Notice that unlike off-path compensation, starting from any constrained optimal toll, this solution leads to worse equilibrium outcomes *and* lower revenue and/or higher deficit.

Parameter Selection The key parameters are based on the NYC tolling program:

Peak-Hour Tolls The $15 baseline toll reflects the original proposal for New York City's congestion pricing program. Toll reduction to $9 was selected based on NYC's revised proposal announced in November 2024, which was adjusted in response to opposition. These provide two focal benchmarks that respondents can evaluate and allow us to compare our contingent compensation approach against a recently policy compromise that was actually implemented.

Projected Traffic Reduction The expected traffic reduction with a $15 toll is taken from the MTA's environmental assessment of the congestion pricing program.[26] The 8% threshold for compensation represents approximately half the projected reduction.

25. The mean monthly trips for respondents who work downtown ($N = 184$) is 8.825 (median = 6) compared to the rest of the sample ($N = 1863$) whose mean monthly trips is 2.055 (median = 0.5), yielding a difference in means of 6.787 (SE = 0.361). Not only is workplace location strongly predictive of driving to downtown, conditioning on other demographics adds very explanatory power: regressing monthly trips on workplace location alone yields adjusted $R^2 = 0.178$; adding a rich set of demographics (income, education, employment, political affiliation, marital status, religion etc.) yields adjusted $R^2 = 0.1931$.

26. The June 2024 Environmental Assesment of the MTA for NYC predicted a 17% reduction in traffic volume. See https://www.mta.info/document/142711.

Monthly Compensation The $120 monthly compensation amount was calculated to ensure approximate budget balance when the compensation is triggered. The calculation takes into account average daily revenue from toll; administrative costs and exemptions (estimated at 50% of gross revenue), and the ratio between total commuters and the number of workers who commute to work in the downtown area. Further details are in Online Appendix E.

4.2.4 Second Stage (belief)

After stating support for the modified policy, respondents answer a second question about their congestion reduction beliefs, allowing us to identify a treatment effect on beliefs. Those who answered a question about NYC in the first stage now answer about their own commute, and those who answer about their own commute at baseline answer about NYC. In other words, each respondent answers both own commute and NYC traffic reduction, in randomized order (before vs. after exposure to treatment).

4.3 Predictions

The model makes three main predictions. First, it predicts a non-trivial amount of equilibrium neglect, manifested in overly pessimistic beliefs about the effect of congestion pricing on commute time. Second, those pessimistic beliefs should in turn be associated with lower support for the policy, which I measure as net support. This interpretation assumes that elicited beliefs capture meaningful variation in respondents' beliefs about traffic rather than expressive restatements or rationalizations of their attitudes. Furthermore, belief must not simply reflect other unobservable dispositions that predict support (e.g., general skepticism or trust in government).

Third, contingent compensation is predicted to increase public support. Notice that absent altruistic motives, contingent compensation should only affect the support of potential recipients, i.e., those who work downtown. Therefore, we should expect the treatment effect to be concentrated in this sub-population. As discussed before, the toll-reduction condition is a more conventional tool that serves as a benchmark. Finally, while the theory does not predict that the treatment will change beliefs, this is a plausible psychological mechanism: providing off-path compensation may signal confidence in the effectiveness of the policy and therefore may make people more optimistic about its congestion-reducing effect. My design allows for testing of belief shifts as a mechanism for increase in support.

4.4 Results

4.4.1 Survey Administration and Sample

The survey was fielded by NORC's AmeriSpeak® Panel in April 2025.[27] The sample consists of a total of 2,056 adults residing in one of the following Metropolitan Statistical Areas: Boston–Cambridge–Newton, MA-NH ($N = 161$); Chicago–Naperville–Elgin, IL-IN ($N = 590$); Los Angeles-Long Beach-Anaheim, CA ($N = 613$); Washington-Arlington-Alexandria, DC-VA-MD-WV ($N = 209$); Dallas-Fort Worth-Arlington, TX ($N = 211$); and San Francisco-Oakland-Berkeley, CA ($N = 272$).[28]

4.4.2 First-Stage Support

This section analyzes baseline opposition for congestion pricing. The next sections quantify the degree of equilibrium neglect—skepticism about travel time reduction—and show that such skepticism strongly predicts opposition.

Figure 4 plots the share of each support stance, with and without sampling weights. An overwhelming majority opposes congestion pricing policy in my sample: 63.2% of the population oppose congestion pricing (43.9% strongly oppose), while only 20.7% support it; the remaining 13% are neutral (choosing "neither support nor oppose"). This translates to -42.5 points net support. Results are qualitatively similar using unweighted data: 66.8% oppose vs. 20.5% support, yielding -46.3 net support.

From this point on I focus on *net support* as a unified single-number, easy-to-interpret measure of support in the population. Appendix Figure F.1 shows net support in each of the six metro areas in the sample. It shows that while there is substantial variation, from -57.1% in Dallas to -20.8% in San Francisco, congestion pricing remains largely unpopular across these different regions. For the rest of the section, I report unweighted results, which have more straightforward interpretation. As discussed below, applying sampling weights does not meaningfully change my main results.

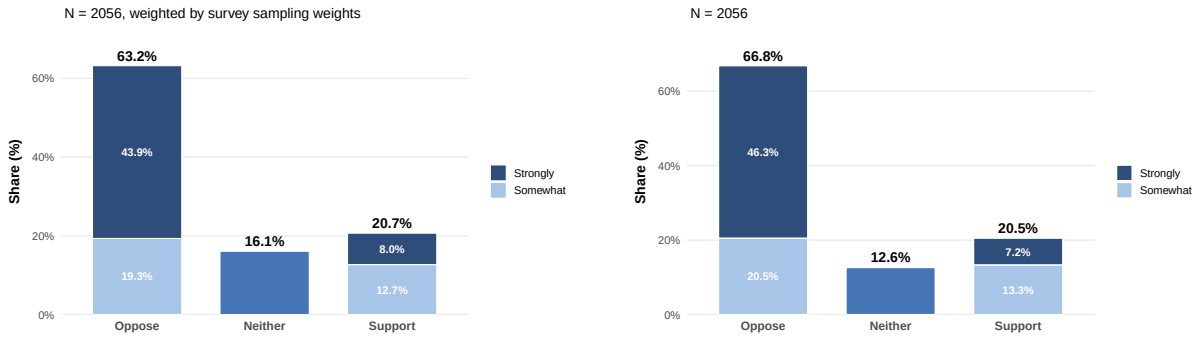
4.4.3 Beliefs About Own Commute Times

In the context of congestion pricing, equilibrium neglect is over-pessimism about congestion reduction. I quantify the degree of EN in the data by comparing respondents'

27. AmeriSpeak® is a probability-based, nationally representative household panel operated by NORC at the University of Chicago. Households are randomly selected using area probability and address-based sampling from the NORC National Sample Frame, covering approximately 97% of the U.S. household population, excluding only P.O. Box addresses, some unlisted USPS addresses, and certain new constructions. The weighted AAPOR RR31 response rate for this panel is 26.1%.

28. These six MSAs were selected based on NORC's feasibility constraints, and in order to represent reasonably diverse geographic regions across the United States that suffer from high congestion. The 2024 INRIX Global Traffic Scorecard ranks Chicago, Los Angeles, and Boston as 2nd, 3rd, and 4th respectively for U.S. traffic congestion, with Washington DC (9th), San Francisco (15th), and Dallas (20th) also appearing in the top 20 (INRIX and Pishue 2024).

Figure 4: Distribution of Support Stance



(a) Weighted (b) Unweighted

Note: Distribution of support for $15 congestion toll. Panel (a) shows weighted results using survey sampling weights. Panel (b) shows unweighted results.

beliefs about traffic reduction to an objective measure of time savings. I start with predictions about changes to own commute time. For each zip code z in my sample, I model post-congestion pricing time savings on trips from z to downtown as:

$$\text{time saving}_z = \rho_z(\text{current rush-hour time}_z - \text{current off-peak time}_z), \tag{3}$$

where $\rho_z = 1$ represents complete elimination of congestion (such that rush-hour travel time equals off-peak time) and $\rho_z = 0$ represents no improvement. This formulation allows natural extrapolation across locations accounting for the fact that tolling can only improve travel times to the extent congestion was slowing traffic to begin with.

Assuming $\rho_z = \rho$ across locations, I estimate ρ from NYC data by plugging pre- and post-congestion pricing travel times from Moshes and Moshes (2025) into the formula above.[29] Defining rush hours as the tolling period in NYC (5 am – 9 pm) and off-peak as 1 am – 3 am, I obtain $\hat{\rho} = 0.566$, i.e., on average, 56.6% of rush-hour congestion delay was eliminated.[30] Further details are in Appendix Table F.1.

I estimate time savings for each respondent by using Google Maps Distance Matrix API to calculate commute times from their zip codes to downtown during both rush and off-peak hours. Multiplying the difference by $\hat{\rho}$ yields *estimated* time savings for each respondent, which I contrast with their *perceived* time savings from survey responses.

The distribution of calculated time savings (Figure 5; orange bars) shows that most

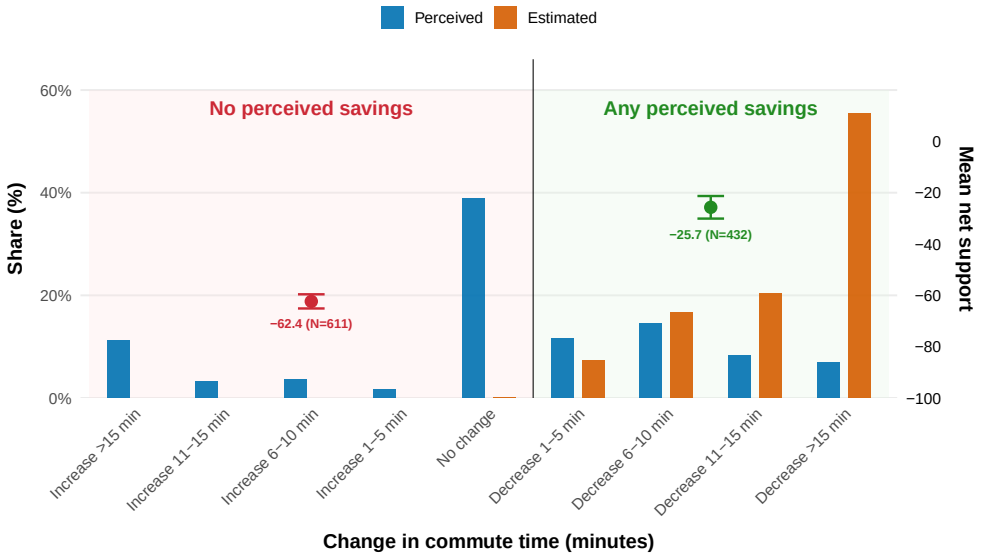
29. I am grateful to Benjamin and Joshua Moshes for giving me access to their raw data.

30. Since the routes leading to downtown in Moshes and Moshes (2025) only include the entrances, extrapolating $\hat{\rho}$ from this data to my sample implicitly assumes that much of the delay people experience happens in entering the city. If commute times on other parts of the route improve by less, this approach may overstate time savings for people's typical commutes. On the other hand, it also likely *understates* predicted time savings under a $15 toll, since the toll in NYC is $9. Either way, it provides a useful benchmark for measuring the degree to which respondents' commute times could improve under congestion pricing.

respondents currently suffer from significant traffic delays and are therefore estimated to experience significant time reductions, with the largest share (55%) estimated to save more than 15 minutes per trip. In contrast, perceived time savings (blue bars) is consistent with strong equilibrium neglect among a solid majority of respondents: About 39% believe congestion pricing would result in no change at all to travel times, and another 20% believe travel times would *increase*. Less than 30% believe that their travel time would decrease by 6 minutes or more.

The 39% share of respondents selecting the option "Driving time would not change" is a lower bound on EN among respondents which does not rely on my measure of estimated time savings (these respondents under-estimate any positive time saving). Notably, this share is also unlikely to be driven by the choice of the scale (15-minute increase to 15-minute decrease) or misunderstanding (e.g., mistaking "increase" for "decrease"). Interestingly, these beliefs appear to reflect substantive views rather than cognitive noise and confusion: limiting the sample to respondents with below-median cognitive uncertainty (i.e., those reporting certainty in their prediction $> 50\%$; $N = 460$), beliefs are *more* pessimistic, with about 51% of respondents predicting no change, and another 20% predicting an increase in travel time.[31]

Figure 5: Estimated Time vs. Perceived Time Saving and Net Support



Note: Perceived vs. estimated time savings from congestion pricing ($N = 1{,}043$). Orange bars show estimated savings, calculated as $\hat{\rho} \times$ (rush-hour time − off-peak time) with $\hat{\rho} = 0.566$ is the mean reduction estimated from Moshes and Moshes (2025) (see Table F.1). ZIP-code level travel times are obtained from Google Maps Distance Matrix API. Blue bars show perceived savings from survey responses. Dots indicate mean net support by perceived savings group ("No perceived savings" in red vs. "Any perceived savings" in green), with error bars showing standard errors.

31. In light of the discussion in Section 3.2.1, the fact that conginitive uncertainty is negatively correlated with attenuated beliefs may also suggest that in the context of congestion pricing, equilibrium skepticism, rather than equilibrium naïveté, is the more prevalent form of equilibrium neglect.

Beliefs are consistent with a large share of respondents under-appreciating equilibrium effects, with many estimating them at approximately zero. How much does this explain variation in policy support? The dots in Figure 5 split the sample by whether respondents believe congestion pricing will reduce travel time (green dot; $N = 432$) or not (red dot; $N = 611$), showing a 36.7 percentage point difference in net support ($p < 0.001$). Table 1 regresses net support on an indicator for believing in any travel time reduction. Column 1 replicates the result from Figure 5 and shows that this belief indicator explains about 5 percent of the variation in support. Adding demographic controls (Column 2) increases adjusted R^2 from 0.049 to 0.134, implying beliefs alone explain about 40% as much variation as the model that includes a rich set of demographics.

Elicited beliefs are inherently noisy, potentially attenuating the relationship between beliefs and support. Columns 3–4 restrict the sample to respondents with below-median cognitive uncertainty, and show that among these respondents, beliefs explain 18.9% of the variation. Furthermore, following the discussion from the theory section (Section 3.4.2), we should expect beliefs to matter more for individuals who (i) drive more to the tolled area, and (ii) have a higher value of time. Columns 5–6 further restrict the sample to respondents with above-median downtown driving frequency,[32] which increases the R^2 to 0.223. To proxy for high value of time, columns 7–8 additionally limit the sample to respondents with above-median income, which further increases the R^2 to 0.308. Looking at adjusted R^2 with and without controls, demographics seem to add only modestly to fit, with beliefs alone explaining about 70%-80% of the variation explained by the model that includes full demographic controls.

In summary, while pessimistic beliefs are not the only explanation for the opposition to congestion pricing, they explain a meaningful share of the variation, especially relative to other observables. Moreover, more confident beliefs are stronger predictors of support, particularly among subsamples that stand to gain more from reduced commute times.

Perceived Benefit To map these results more directly to the theory, I construct a measure of actual and perceived gain from congestion pricing and correlate it with support. The measure converts time savings into dollar terms. The per-trip perceived benefit index for respondent n is given by:

$$\tilde{\delta}_n^{trip} = \theta_n \times \text{perceived time savings}_n - \$15, \tag{4}$$

where θ_n represents the value of time (VOT), set at 50% of hourly household income following standard practice (Belenky 2011), and perceived time saving is the belief about

32. The response options are: "Not applicable (I don't own or have access to a car)," "Never," "Less than once a month," "1-3 times per month," "1-2 times per week," "3-4 times per week," "5 or more times per week." The median response is "less than once a month."

Table 1: Perceived Savings and Net Support

	Whole Sample		CU		CU & Trips		CU & Trips & Income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-62.357		-74.924		-79.710		-80.328	
	(3.182)		(3.981)		(6.088)		(9.190)	
Any Perceived Savings	36.662	28.204	76.427	68.840	83.043	77.715	102.550	116.403
	(4.944)	(5.151)	(7.404)	(7.900)	(11.060)	(13.233)	(16.591)	(23.666)
Full Demographics		✓		✓		✓		✓
R^2	0.050	0.181	0.189	0.349	0.223	0.476	0.308	0.721
Adjusted R^2	0.049	0.134	0.187	0.259	0.219	0.277	0.300	0.379
Observations	1043	1030	460	455	198	197	88	88

Note: The table reports results from regressing net support for congestion pricing on an indicator that equals one if the respondent believes congestion pricing will reduce travel time. Columns 1–2 include the whole sample. Columns 3–4 restrict the sample to respondents with below-median cognitive uncertainty. Columns 5–6 further restrict to respondents with above-median self-reported driving trips downtown. Columns 7–8 additionally restrict to respondents with above-median income. Demographic controls include income, sex, age, ethnicity, education, metropolitan area, housing type, home ownership, employment status, marital status, partisanship, ideology, trust in media, religion and religious attendance, household size, and phone ownership. Standard errors are in parentheses.

respondents' own commute.[33] Taking into account trip frequency, $\tilde{\delta}_n^{month}$ multiplies per-trip benefit by the self-reported number of monthly driving trips.

Figure 6 plots net support against these measures. Panels a and c plot net support against perceived per-trip and monthly benefit. An additional perceived per-trip dollar benefit is associated with a 2.033 points increase in net support (SE = 0.326). At the monthly level, additional $10 increase in perceived benefit are associated with an increase of 1.3 points in net support (SE = 0.044). Both relationships are highly statistically significant. Panels b and d plot net support against *estimated* benefits. They show that, unsurprisingly, estimated benefit measures are essentially uncorrelated with support, with significantly smaller coefficients, both not different from zero. Therefore, living in congestion-prone areas that would typically benefit from the time saving associated with congestion pricing is not associated with higher support congestion pricing.

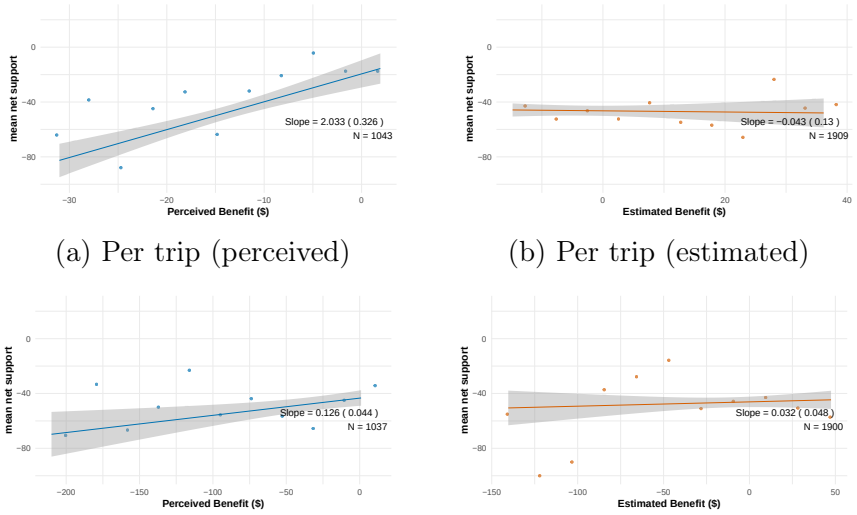
4.4.4 Beliefs about NYC Commute Times

While predictions about one's own time savings are likely the most relevant measure for determining attitudes toward congestion pricing, they are subjective, subject to interpretation, and hypothetical. As an alternative measure, half of the respondents were asked to estimate the change in commute time through Holland Tunnel after the implementation of congestion pricing in NYC. This measure has concrete interpretation and can be easily compared to objective and verifiable ground truth.

Figure 7 plots the distribution of responses to the question about NYC. While less

33. This index maps to $\delta_n(T, T^0; \eta)$ from the theory section with the subtle difference that rather than plugging in the true effect and discounting by η, I directly plug in a potentially attenuated effect.

Figure 6: Benefit and Net Support



(a) Per trip (perceived)

(b) Per trip (estimated)

(c) Monthly (perceived)

(d) Monthly (estimated)

Note: Estimated and perceived benefits from congestion pricing are compared to net support. *Perceived benefit* (panels a and c) is calculated as $\theta_n \times$ perceived time savings – \$15, where θ_n is the value of time, estimated at 50% of hourly household income, and perceived time savings are taken from survey responses. *Estimated benefit* (panels b and d) replaces perceived time savings with estimated time savings (see Figure 4). Monthly benefit (panels c and d) multiplies the relevant benefit by the self-reported number of driving trips to downtown. Bins are equally spaced, and the fitted line is from a linear regression of net support on the relevant benefit measure. Benefit measures are winsorized at the 5th and 95th percentiles.

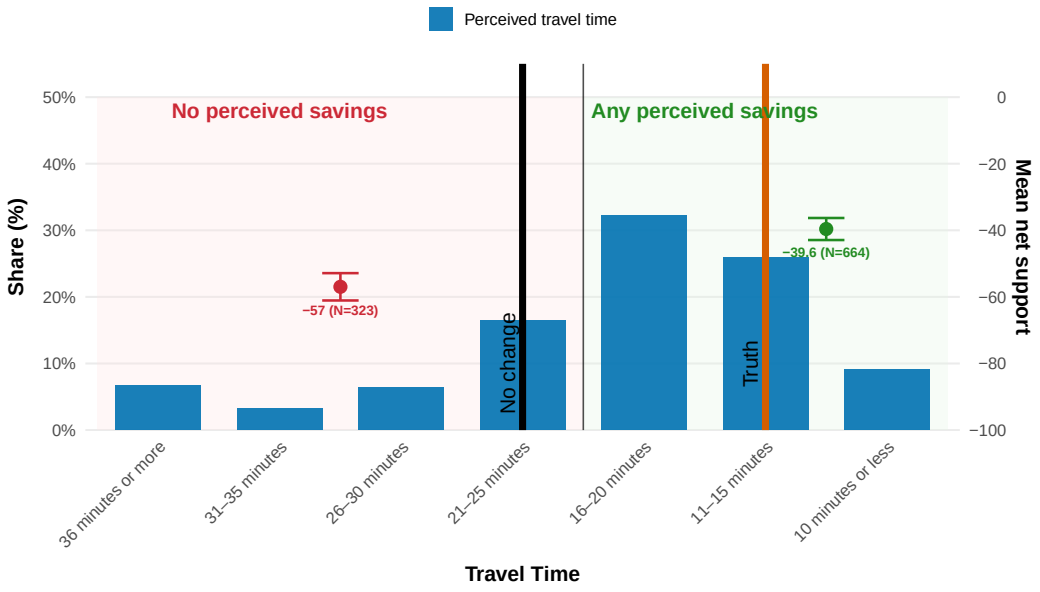
prevalent than own-time predictions, these beliefs still show strong skepticism towards congestion-reducing effects of tolls: the actual post-congestion pricing travel time through Holland tunnel was around 11 minutes, yet 65% of respondents estimate it to be 16 minutes or more. Notably, 33% of respondents estimate it at 21 minutes or more, indicating little to no improvement, or even a worsening in traffic conditions compared to the 23 minutes baseline.

Beliefs about NYC are also predictive of support for congestion pricing: Figure 7 shows that the difference in net support between respondents indicating a belief in any meaningful improvements ($N = 664$) and those who believe that travel times remained roughly the same or worse ($N = 323$) is 16.4 points ($p = 0.001$). However, as expected, these beliefs are less predictive than beliefs about one's own commute times (shown in Figure 5).[34]

To summarize, I find substantial equilibrium neglect in the sample, reflected in pes-

34. Measuring EN using beliefs about NYC likely attenuates both the prevalence of equilibrium neglect and the extent to which it predicts a person's stance on congestion pricing. The first question about support contains information about "other cities" as part of supporters' arguments ("Supporters cite research from other cities suggesting it can cut traffic by about 15–20% and reduce travel times by 10–15 minutes"), potentially mechanically shrinking the prevalence of respondents who believe in no improvement. Additionally, if beliefs about NYC traffic are a noisy proxy for beliefs about one's own commute (because, e.g., they view NYC as fundamentally different than their own city), the correlation between beliefs and policy attitudes is attenuated.

Figure 7: Perceived Travel Time in NYC and Net Support



Note: Perceived travel times through the Holland Tunnel under congestion pricing ($N = 987$). Blue bars show the distribution of beliefs about the new travel time. The solid black line ("no change") marks no meaningful change relative to pre-congestion pricing travel time. The solid orange line ("Truth") indicates the estimated travel time based on Moshes and Moshes (2025). Dots indicate mean net support by perceived savings group ("No perceived savings" in red vs. "Any perceived savings" in green), with error bars showing standard errors.

simistic beliefs about congestion reduction—both in hypothetical estimates about respondents' own commute time reductions, and in their estimate of its actual effect in NYC. These pessimistic beliefs are in turn correlated with opposition to the policy.

4.4.5 Second Stage: Main Results

Having shown that pessimism is prevalent and predicts opposition to congestion pricing, I now turn to the remedy: are policies that include contingent compensation indeed robust to equilibrium neglect? We should expect the effect of EN-robustness to be concentrated in downtown workers, who are the potential recipients of the compensation. Other respondents should only respond to the treatment to the extent that they are altruistic towards people who work in the downtown area or have other fairness concerns.

Ideally, the treatment would be stratified and block-randomized based on baseline support to ensure balanced baseline support across conditions. Unfortunately, due to technical limitations, the vendor was unable to implement this stratification. This resulted in an imbalanced treatment allocation where baseline net support in Control is 6.3 points higher than in EN-robust and Toll Reduction, likely due to chance sampling variation (p-value from a regression testing difference in means across treatments = 0.08). The two-stage design of the experiment allows me to correct for this pre-treatment imbalance by differencing out baseline support (as in Figure 8 below) or by controlling for it (as in Table 2; this is the specification I pre-registered).

Panel a in Figure 8 compares net support among downtown workers in EN-robust and Control across stages, using a difference in differences design. In EN-robust, second-stage net support increases 52.2 points (from −49.3% to +2.9), effectively neutralizing the initially widespread opposition. In Control there is a much smaller 3.9 points increase (from −26 to −22), resulting in a diff-in-diff estimate of 48.3 (SE = 12.208).

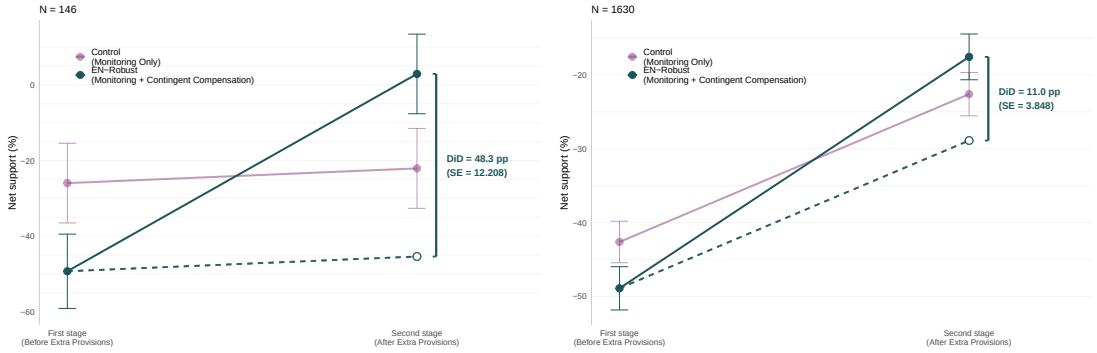
Looking at the whole sample (Figure 8, panel (b)), support increases from baseline to second stage in both treatments, but this increase is 11 points higher in EN-robust than in Control (SE = 3.848). The increase in Control likely reflects the effect of monitoring alone and, potentially, the effect of repeated elicitation.[35]

The first row in Table 2 shows similar results from a regression of second-stage net support on treatment, controlling for baseline support, which remain stable with the inclusion of a rich set of controls. Net support under EN-robust is 38.545–42.351 points higher than control among downtown workers (SE = 11.255–14.028) and 9.127–10.154 points higher in the whole sample (SE = 3.378–3.411). Taken together, these results show that contingent compensation indeed increases support for congestion pricing in the population, and, as expected, its effects are concentrated, and are particularly large,

35. More speculatively, some respondents might have interpreted "monitoring" to mean that the policy would be scrapped if the evidence suggested it was ineffective, which could have resulted in increased support.

within the group it targets. Online Appendix Table F.2 replicates these regressions weighting observations by sampling weights. The coefficient on EN-robust for the whole sample slightly decreases but remains statistically different from zero (7.176–8.369 net support points; SE = 3.401–3.499).

Figure 8: Effect of Treatment on Support for Congestion Pricing



(a) Downtown Workers (b) Whole Sample

Note: Net support across stages in EN-robust and Control. Panel (a) plots the results for downtown workers only ($N = 146$). Panel (b) reports results for the whole sample ($N = 1630$). Each figure plots mean net support in the first and second stage. The solid teal line represents EN-robust, and the solid purple line represents Control. Error bars indicate standard errors. The dashed teal line shows the counterfactual of EN-robust under parallel trends, based on the observed change in Control. The bracket indicates the difference-in-differences estimate, reported in net support percentage points.

Table 2: Treatment Effects on Net Support

	Downtown Workers			Whole Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
EN-robust	40.908	38.545	41.642	9.244	9.127	9.777
	(11.255)	(11.773)	(14.019)	(3.379)	(3.378)	(3.434)
Toll Reduction	4.783	1.309	-2.026	-8.564	-8.210	-6.773
	(13.380)	(13.732)	(15.416)	(4.040)	(4.039)	(4.059)
Baseline Support	✓	✓	✓	✓	✓	✓
Basic Demographics		✓	✓		✓	✓
All Demographics			✓			✓
R^2	0.458	0.473	0.641	0.375	0.380	0.405
Observations	184	184	181	2043	2043	2021

Note: The table reports results from regressing second-stage net support congestion on treatment, controlling for baseline support; the omitted category is Control. Columns 1–3 restrict the sample to respondents who work in the downtown area, and are eligible for contingent compensation. Columns 4–6 include the whole sample. "Basic demographics" control for income, sex, age, ethnicity and education. "All demographics" additionally include metropolitan area, housing type, home ownership, employment status, marital status, partisanship, ideology, trust in media, religion and religious attendance, household size, and phone ownership. Standard errors are in parentheses.

Toll Reduction To benchmark the effect, I included a treatment reducing the toll from \$15 to \$9. The second row in Table 2 shows that its effects are significantly smaller. It is positive but statistically indistinguishable from zero among downtown workers, and if anything, has a *negative* effect on the whole population. As treatment effect is measured relative to Control, this suggests that the effect of reducing the toll is similar, or smaller in magnitude to a simple accountability treatment. A potential explanation could be that EN-driven skepticism towards the policy may make respondents positively react to increased accountability.[36]

4.4.6 Additional Analyses

Treatment Effects by Response Time Apparent opinion shifts in surveys may be driven by methodological artifacts rather than genuine reaction to the substance. Finding that contingent compensation increases support especially among its intended recipients is consistent with treatment effects that are indeed driven by reaction to the actual content of the policy rather than other survey design artifacts. Appendix Table F.3 shows that treatment effect on respondents who spend above the median time (31 seconds) on the second-stage support question is 17.24–18.18 points higher (SE = 6.94–6.99) than those below the median. The treatment effect on the below-median group is statistically indistinguishable from zero. This suggests that the contingent compensation mechanism affects only respondents who spent enough time to understand its content.[37]

Treatment Effect on Beliefs While the model makes no clear prediction about treatment effects on beliefs, contingent compensation might plausibly work through a signaling mechanism—policymakers committing to large payments if traffic does not fall could increase citizens' confidence in the equilibrium traffic-reduction prediction. Finding that effects are concentrated among targeted recipients suggests that this is unlikely to be the case. Indeed, as Online Appendix Table F.4 shows, regressing second-stage beliefs—either about the respondent's own commute or about effects in NYC—on treatment yields small coefficients that are all statistically indistinguishable from zero (-0.69 to -0.02; SE = 0.59–0.61).

36. In addition, finding that toll reduction has a small effect on support may not be all that surprising from a theoretical point of view: an equilibrium-neglecting agent still perceives negative benefits from a \$9 toll, and should still oppose it. At the same time, agents with perfect foresight should only change their mind about the toll if they believe the reduction in toll is worth more than the potential reduction in time savings it will cause. Therefore, the overall effect of a toll reduction is ambiguous.

37. Splitting the sample by whether respondents spent more than 18 seconds on the question—the minimal time required for average human reading speed (238 words/minute) to process the 71 words in the added EN-robust clause—yields similar results: interaction effects of 18.19–18.6 (SE = 8.24–8.27) and main effects indistinguishable from zero (-1.08 to -5.63; SE = 6.94–7.31). In the absence of direct attention checks, one can also think of these (non-preregistered) specifications as a form of screening less attentive respondents.

5 Congestion Pricing: Policymakers

The previous section documents equilibrium neglect among the general public in the context of congestion pricing, showing that it is prevalent and predicts policy opposition, while adding a contingent compensation to the policy leads to higher public support. To understand how these findings shape political feasibility, I turn to the supply side. In the absence of referenda, political feasibility is ultimately driven by the willingness of policymakers—both civil servants and elected officials—to engage with a policy and implement it. This willingness may depend on their own assessment of policy effectiveness and their perception of public support.

This raises several questions: Are policymakers themselves prone to equilibrium neglect? Do they perceive the public to be equilibrium neglectful? How do these beliefs map to policymakers' stance on congestion pricing? This section presents results from surveys eliciting both first- and second-order beliefs among two groups of policymakers: US state legislators and civil servants from the US, UK, and Australia.

5.1 Sample

State Legislators To survey politicians, I rented a booth at the 2025 National Convention of State Legislators Summit and invited state legislator attendees to participate in a short in-person survey. Legislators were identifiable by a special badge color. Participants were offered compensation ($5-25 Starbucks gift cards, randomly drawn) with the option to donate it to a food bank in their state. Over two days, 130 participants started the survey, which resulted in 117-118 responses to relevant questions after attrition.

Civil Servants I surveyed 58 federal-level civil servants from the US, UK, and Australia recruited through the Policymakers Lab, a research initiative that connects policymakers worldwide with academics to study how evidence is used in decision-making. Participants received $15 Amazon gift cards (or its equivalent amount in their local currency).

5.2 Design

The state legislators survey was conducted in collaboration with another researcher studying housing development support. The order of the parts of the survey was randomized. Both the state legislators and civil servants surveys included additional elicitation of EN and perceived public EN on other topics, to be discussed in future iterations of this paper. Here I focus on congestion pricing questions that paralleled the survey experiment.

All participants first see a congestion pricing policy identical to Control from the general-public survey (i.e., the baseline policy with the added monitoring clause). Rather than asking about their own commute time, participants are asked to predict the change

in commute time for a driver who currently spends 30 minutes on his commute every morning, compared to the 20 minutes this same route takes without congestion. The possible answers have a similar symmetric structure as the survey experiment: more than 10 minute decrease, 3–10 minute decrease, no meaningful change, 3–10 minute increase, more than 10 minute increase. Among the possible answers, 3–10 is the only economically reasonable one, with the rest representing an under- or over-appreciation of the time-saving effects of congestion pricing.[38]

Participants are then asked to predict the share of the general public who would mention any time reduction when asked to predict the consequences of the policy. This question captures policymakers' perceived level of equilibrium neglect among the public, focusing on what I view as the most ecologically relevant aspect: whether the equilibrium effect is top of mind for citizens when faced with such a policy. Next, participants indicate their own level of support[39] followed by a question asking them to predict the level of support among the public. Finally, all participants then see the EN-robust version of the policy, with the added contingent compensation highlighted,[40] and indicate again their own level of support and then estimate the level of public support for that version of the policy.

5.3 Results

Beliefs I start with first-order beliefs. State legislators exhibit an under-appreciation of time savings from congestion pricing similar to that of the general public. As shown in Figure 9, 43% predict no improvement in commute times. This suggests that EN may extend from voters to elected officials. In contrast, only 15% of civil servants predict no improvement; a large majority correctly anticipate time savings, with the largest share (78%) choosing the "textbook" 3-10 minute decrease. These beliefs strongly predict support: among state legislators, net support is 37 points higher for those who predict time savings than for those who do not. The difference is larger among civil servants, at 140 points, though this should be interpreted with caution, as the no-time-savings group in this sample contains only nine respondents.

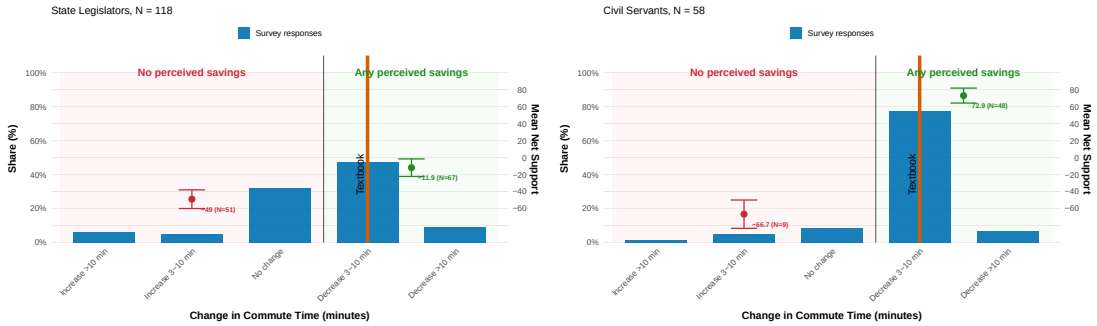
State legislators and civil servants both expect equilibrium effects to be less salient to

38. Predicting a decrease larger than 10 minutes likely over-estimates the reduction, as the question mentions 20 minutes as the free-flow travel time. I interpret choosing "no meaningful change" as representing under-appreciation of equilibrium effects. While it could also be interpreted as reflecting a genuine belief in a meaningful time reduction of one or two minutes, results cited above from NYC suggest this would likely underestimate the effect on travel times, given NYC implemented a 40% lower toll and saw 60% improvement relative to free-flow traffic.

39. To emphasize that the question asks about their attitude in their capacity as policymakers, the question reads: "How much would you recommend or support a similar policy in your current role?" with the same five-point scale as in the general-public survey.

40. Since participants view both the Control and EN-robust versions of congestion pricing from the general-public survey, this effectively serves as a within-subject replication of that survey's second stage.

Figure 9: Policymakers' Perceived Time Savings



(a) State Legislators (b) Civil Servants

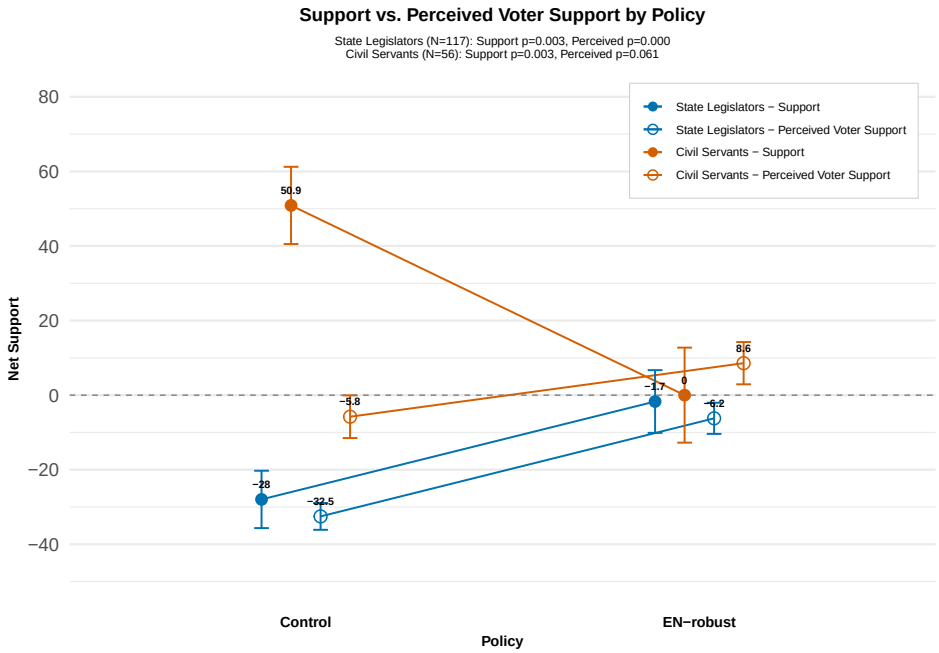
Note: Perceived time savings from congestion pricing for a hypothetical driver. The baseline time in the question is 30, and the free-flow time is 20 minutes. Blue bars show the distribution of beliefs about time savings. The solid orange line ("Textbook") indicates meaningful traffic reduction up to the free-flow speed. Dots indicate mean net support by perceived savings group ("No perceived savings" in red vs. "Any perceived savings" in green), with error bars showing standard errors.

the general public. Figure G.1 shows they predict respondents from the general public are 23 percentage points more likely to mention congestion pricing's direct effect (increased costs) than its equilibrium effect (shorter commute). These second-order beliefs also strongly predict policy support. Table G.1 shows that a one percentage-point increase in the estimated share of respondents from the public who mention time savings is associated with a 0.93 (SE = 0.28) points increase in a state legislator's net support and a 2.03 (SE = 0.42) points increase for a civil servant. These results hold when controlling for the participants' own beliefs, suggesting that a perception of public understanding of a policy is an important input in policymakers' support even conditional on their own assessment of the policy.

Support for EN-robust Policies Finally, I look at support for Control and EN-robust congestion pricing, and perceived public support for each. Figure 10 plots net support and perceived public support (converted to net support points) in both groups. State legislators' own support for EN robust is more than 26 points higher than Control, moving from -28 to -1.7. Their perception of public support follows a similar pattern, with estimated public net support rising from -32.5 to -6.2 between Control and EN-robust.

For civil servants, the pattern is very different. While support for the standard congestion pricing policy (Control) is overwhelmingly high (+51 net support), it drops significantly for the EN-robust version to an equal split (net support = 0). At the same time, civil servants on average seem to predict that public support for the EN-robust congestion pricing is higher, with estimated mean public net support increasing 14 points (-5.8 to $+8.6$). Interestingly, despite their different starting points, both elected officials and civil servants converge on a similar, roughly neutral level of support for the EN-robust

Figure 10: Policymakers: Net Support and Perceived Net Support for Control and EN-robust



Note: Net support and perceived public net support for Control and EN-robust. The full circles represent actual net support for each policy, and the hollow circles represent perceived public net support. Error bars indicate standard errors. Orange: civil servants ($N = 56$). Blue: state legislators ($N = 117$). p-values: paired t-tests of the null of equal means.

congestion pricing policy.

5.4 Discussion

Taken together, these findings can offer two insights. First, the results highlight the potential of EN-robust policies for easing opposition from both elected officials and the public, even if it comes at the cost of less enthusiastic support from administrators. I speculate that explaining outcome and deficit equivalence will likely be effective in convincing civil servants, especially when combined with evidence on increased public support, but this remains an open question for future research. Second, these findings may help explain why contingent compensation mechanisms are not currently common in public policy: those tasked with designing policies—civil servants—seem to prefer conventional tools. This preference, however, may lead them to champion policies whose political feasibility is questionable.

6 Conclusion

This paper describes a problem and proposes a solution. At its core, the problem is one any economist who has picked up a newspaper will recognize: the bread and butter of the

discipline—thinking through the implications of best response—is a mode of reasoning that does not always come naturally to the general public. Similarly, at its core, the remedy is hardly exotic: if you want to make people happier with a policy, simply offer them more money. The key to a remedy, I argue, is recognizing that the same cognitive bias that can limit the set of policies we are able to choose from also creates scope for "policy arbitrage," allowing us to generate perceived gains at no fiscal cost, and align public support with the policy's actual benefits.

While this paper focuses on congestion pricing, equilibrium neglect is a general phenomenon that likely constrains the political feasibility of a wide range of socially beneficial policies, from carbon taxation to financial regulation. And while off-path contingent compensation is designed to be similarly portable, tailoring it to a particular context requires carefully thinking about who the potential spurious opposition is, and what to condition on in order to appease it.

Future research could explore several avenues. First, applying and testing this framework in other policy domains is a natural next step. Second, the disconnect identified between civil servants and elected officials merits further investigation: what makes policy experts prefer conventional tools, and how can they be persuaded of their utility in bridging the gap between optimal policy and public opinion?

Finally, contingent compensation mechanisms may extend beyond the bottleneck this paper is focused on: they could address opposition driven by cognitive biases other than equilibrium neglect, and they could also provide insurance against outcomes in settings with real uncertainty, creating a neutral and transparent tool that may help bridge the gap between what is optimal and what is politically feasible.

References

Allcott, Hunt, and Christopher Knittel. 2019. "Are Consumers Poorly Informed about Fuel Economy? Evidence from Two Experiments." *American Economic Journal: Economic Policy* 11 (1): 1–37.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky. 2019. "Regressive Sin Taxes, with an Application to the Optimal Soda Tax." *The Quarterly Journal of Economics* 134 (3): 1557–1626.

Amador, Manuel, and Javier Bianchi. 2024. "Bank Runs, Fragility, and Credit Easing." *American Economic Review* 114 (7): 2073–2110.

American Association for Public Opinion Research. 2023. *Question Wording.* Technical report. American Association for Public Opinion Research. https://aapor.org/wp-content/uploads/2023/01/Question-Wording.pdf.

Andersson, David, and Jonas Nässén. 2016. "The Gothenburg congestion charge scheme: A pre–post analysis of commuting behavior and travel satisfaction." *Journal of Transport Geography* 52:82–89.

Baranzini, Andrea, Stefano Carattini, and Linda Tesauro. 2021. "Designing Effective and Acceptable Road Pricing Schemes: Evidence from the Geneva Congestion Charge." *Environmental and Resource Economics* 79 (3): 417–482.

Belenky, Peter. 2011. *Revised departmental guidance on valuation of travel time in economic analysis.* Technical report. Washington, DC: US Department of Transportation.

Bernheim, B. Douglas, and Dmitry Taubinsky. 2018. "Behavioral Public Economics." In *Handbook of Behavioral Economics: Applications and Foundations 1,* edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 1:381–516. Handbook of Behavioral Economics - Foundations and Applications 1. North-Holland. Accessed September 11, 2025. https://www.sciencedirect.com/science/article/pii/S2352239918300022.

Binetti, Alberto, Francesco Nuzzi, and Stefanie Stantcheva. 2024. *People's Understanding of Inflation.* Working Paper.

Bó, Inácio, and Rustamdjan Hakimov. 2023. "Pick-an-Object Mechanisms." *Management Science,*

Börjesson, Maria, Jonas Eliasson, and Carl Hamilton. 2016. "Why experience changes attitudes to congestion pricing: The case of Gothenburg." *Transportation Research Part A: Policy and Practice* 85:1–16.

Boulding, William, and Amna Kirmani. 1993. "A Consumer-Side Experimental Examination of Signaling Theory: Do Consumers Perceive Warranties as Signals of Quality?" *Journal of Consumer Research* 20 (1): 111–123.

Butler, Daniel M. 2011. "Can Learning Constituency Opinion Affect How Legislators Vote? Results from a Field Experiment." *Quarterly Journal of Political Science* 6 (1): 55–83.

Caplan, Bryan. 2001. "What Makes People Think Like Economists? Evidence on Economic Cognition from the "Survey of Americans and Economists on the Economy"." *The Journal of Law and Economics* 44 (2): 395–426.

Carattini, Stefano, Andrea Baranzini, and Rafael Lalive. 2018. "Is Taxing Waste a Waste of Time? Evidence from a Supreme Court Decision." *Ecological Economics* 148:131–151.

Carattini, Stefano, Maria Carvalho, and Sam Fankhauser. 2018. "Overcoming public resistance to carbon taxes." *WIREs Climate Change* 9 (5): e531.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–1177.

Chinoy, Sahil, Nathan Nunn, Sandra Sequeira, and Stefanie Stantcheva. 2023. *Zero-Sum Thinking and the Roots of U.S. Political Divides.* Working Paper.

Chu, Junhong, and Pradeep K. Chintagunta. 2011. "An Empirical Test of Warranty Theories in the U.S. Computer Server and Automobile Markets." *Journal of Marketing* 75 (2): 75–92.

Cohen, Shani, and Shengwu Li. 2025. *Sequential Cursed Equilibrium.*

Crémer, Jacques, and Richard P. McLean. 1988. "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions." *Econometrica* 56 (6): 1247–1257.

Dal Bó, Ernesto, Pedro Dal Bó, and Erik Eyster. 2018. "The Demand for Bad Policy when Voters Underappreciate Equilibrium Effects." *The Review of Economic Studies* 85 (2): 964–998.

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. 2025. "Fighting Climate Change: International Attitudes toward Climate Policies." *American Economic Review* 115 (4): 1258–1300.

De Clippel, Geoffroy. 2014. "Behavioral Implementation." *American Economic Review* 104 (10): 2975–3002.

De Clippel, Geoffroy, Rene Saran, and Roberto Serrano. 2019. "Level-k Mechanism Design." *The Review of Economic Studies* 86 (3): 1207–1227.

Douenne, Thomas, and Adrien Fabre. 2022. "Yellow Vests, Pessimistic Beliefs, and Carbon Tax Aversion." *American Economic Journal: Economic Policy* 14 (1): 81–110.

Dreyfuss, Bnaya, Ofer Glicksohn, Ori Heffetz, and Assaf Romm. 2025. "Deferred Acceptance with News Utility." *Management Science,*

Duranton, Gilles, and Matthew A. Turner. 2011. "The Fundamental Law of Road Congestion: Evidence from US Cities." *American Economic Review* 101 (6): 2616–2652.

Eliasson, Jonas. 2014. *The Stockholm congestion pricing syndrome: how congestion charges went from unthinkable to uncontroversial.*

Elmendorf, Christopher S., Clayton Nall, and Stan Oklobdzija. 2025. "The Folk Economics of Housing." *Journal of Economic Perspectives* 39 (3): 45–66.

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive Uncertainty*." *The Quarterly Journal of Economics* 138 (4): 2021–2067.

Esponda, Ignacio, and Emanuel Vespa. 2014. "Hypothetical Thinking and Information Extraction in the Laboratory." *American Economic Journal: Microeconomics* 6 (4): 180–202.

Eyster, Erik, and Matthew Rabin. 2005. "Cursed Equilibrium." *Econometrica* 73 (5): 1623–1672.

Gavan, Malachy James, and Antonio Penta. 2025. "Safe Implementation." *Theoretical Economics.*

Gonczarowski, Yannai A., Ori Heffetz, and Clayton Thomas. 2023. *Strategyproofness-Exposing Mechanism Descriptions.*

Goodwin, P. B. 1997. *Solving congestion (when we must not build roads, increase spending, lose votes, damage the economy or harm the environment, and will never find equilibrium).* Report 97/60. London: ESRC Transport Studies Unit, University College London. Accessed September 12, 2025. https://discovery.ucl.ac.uk/id/eprint/1244/.

Gu, Ziyuan, Zhiyuan Liu, Qixiu Cheng, and Meead Saberi. 2018. "Congestion pricing practices and public acceptance: A review of evidence." *Case Studies on Transport Policy* 6 (1): 94–101.

INRIX, and Bob Pishue. 2024. *INRIX 2024 Global Traffic Scorecard.*

Jaensirisak, S., M. Wardman, and A. D. May. 2005. "Explaining Variations in Public Acceptability of Road Pricing Schemes." *Journal of Transport Economics and Policy (JTEP)* 39 (2): 127–154.

Joskow, Paul L., and Catherine D. Wolfram. 2012. "Dynamic Pricing of Electricity." *American Economic Review* 102 (3): 381–385.

Larsen, Odd I. 1995. "The toll cordons in Norway: an overview." *Journal of Transport Geography* 3 (3): 187–197.

Li, Shengwu. 2017. "Obviously Strategy-Proof Mechanisms." *American Economic Review* 107 (11): 3257–87.

Metropolitan Transportation Authority (MTA). 2024a. *Central Business District (CBD) Tolling Program Re-Evaluation.* Technical report. Accessed January 24, 2025. https://mta.info/document/142711.

Metropolitan Transportation Authority (MTA). 2024b. *Central Business District (CBD) Tolling Program Re-Evaluation 2.* Technical report. Accessed January 24, 2025. https://mta.info/document/158191.

Micheletto, Luca, Dylan T. Moore, Daniel Reck, and Joel Slemrod. 2025. "An inverse-Ramsey tax rule." *Journal of Public Economics* 251:105501.

Moshes, Benjamin, and Joshua Moshes. 2025. *Congestion Pricing Tracker.*

Murray, Brian, and Nicholas Rivers. 2015. "British Columbia's revenue-neutral carbon tax: A review of the latest "grand experiment" in environmental policy." *Energy Policy* 86:674–683.

Myerson, Roger B. 1981. "Optimal Auction Design." *Mathematics of Operations Research* 6 (1): 58–73.

Nagel, Lea, and Roberto Saitto. 2025. "A Measure of Complexity for Strategy-Proof Mechanisms." Accessed July 10, 2025.

Niederle, Muriel, and Emanuel Vespa. 2023. "Cognitive Limitations: Failures of Contingent Thinking." *Annual Review of Economics* 15 (Volume 15, 2023): 307–328.

Odeck, James, and Svein Bråthen. 1997. "On public attitudes toward implementation of toll roads—the case of Oslo toll ring." *Transport Policy* 4 (2): 73–83.

Pigou, Arthur Cecil. 1920. *The Economics of Welfare.* 1st ed. London: Macmillan / Co., Limited. ISBN: 978-0-230-24931-8, accessed September 4, 2025. https://link.springer.com/book/9780230249318.

Pycia, Marek, and Peter Troyan. 2023. "A Theory of Simplicity in Games and Mechanism Design." *Econometrica* 91 (4): 1495–1526.

Rees-Jones, Alex, and Dmitry Taubinsky. 2020. "Measuring "Schmeduling"." *The Review of Economic Studies* 87 (5): 2399–2438.

Ronn, Ehud I., and Avinash K. Verma. 1986. "Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model." *The Journal of Finance* 41 (4): 871–895.

Sapienza, Paola, and Luigi Zingales. 2013. "Economic Experts versus Average Americans." *American Economic Review* 103 (3): 636–642.

Schuitema, Geertje, Linda Steg, and Sonja Forward. 2010. "Explaining differences in acceptability before and acceptance after the implementation of a congestion charge in Stockholm." *Transportation Research Part A: Policy and Practice* 44 (2): 99–109.

Schuman, Howard, and Stanley Presser. 1996. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* SAGE. ISBN: 978-0-7619-0359-8.

Semken, Christoph. 2025. "The Marginal Impact of Emission Reductions: Estimates, Beliefs and Behavior." *Working Paper.*

Siena College Research Institute. 2024. *Hochul Job Approval & Favorability Ratings Up a Little, Remain Negative; Only 33% of Voters Would Re-Elect Hochul, 57% Want 'Someone Else' – Siena Research Institute.*

———. 2025. *Hochul – with a Big Lead in a Very Early Democratic Primary Look – Gets Good Grades from Voters on Protecting Constitutional Rights & Ensuring Quality Affordable Healthcare, But Not on Making NY More Affordable – Siena Research Institute.*

Spiegler, Ran. 2020. "Behavioral Implications of Causal Misperceptions." *Annual Review of Economics* 12 (Volume 12, 2020): 81–106.

Stantcheva, Stefanie. 2022. *Understanding of Trade.* Working Paper.

Transport for London. 2004. *Congestion charging: Central London impacts monitoring – Second annual report.* Technical report. London: Transport for London. Accessed October 10, 2025. https://content.tfl.gov.uk/impacts-monitoring-report-2.pdf.

Tretvik, Terje. 2003. "Urban Road Pricing in Norway: Public Acceptability and Travel Behaviour." In *Acceptability of Transport Pricing Strategies,* edited by Jens Schade and Bernhard Schlag, 0. Emerald Group Publishing Limited. ISBN: 978-0-08-044199-3, accessed October 10, 2025. https://doi.org/10.1108/9781786359506-005.

Vickrey, William S. 1969. "Congestion Theory and Transport Investment." *The American Economic Review* 59 (2): 251–260.

Weyl, E. Glen. 2010. "A Price Theory of Multi-sided Platforms." *American Economic Review* 100 (4): 1642–1672.

A Proofs

This section contains the proofs for all Results, Propositions, and Theorems, following the order they appear in the paper.

A.1 Result 1

Proof. Any equilibrium must be a cutoff equilibrium. Let $\theta_1 > \theta_2$. If both types play best responses and type θ_1 weakly prefers public transit, then type θ_2 strictly prefers public transit. Similarly, if type θ_2 weakly prefers driving, then type θ_1 strictly prefers driving.

Any cutoff equilibrium with cutoff θ must satisfy the indifference condition:

$$\theta Q(1 - F(\theta); I) = b + p$$

Since the LHS is strictly increasing in θ (both θ and $Q(1 - F(\theta); I)$ are increasing in θ) and ranges from 0 to ∞, while the RHS equals $b + p > 0$, this equation has a unique solution $\theta^*(T)$.

Therefore, there exists a unique Nash equilibrium $a^*(T)$ where $a_n = 1$ iff $\theta_n \geq \theta^*(T)$. □

A.2 Proposition 1

The proposition is proved by the example that follows it.

A.3 Proposition 2

Proof. The proof is constructive. Fix a feasibility rule f that satisfies monotonicity and consensus. Let $N = [0, 1]$ with Lebesgue measure μ, set $A_n = \{0, 1\}$ for all $n \in N$, and define utilities:

$$v_n(a; T) = a_n(1 - T) - (K + 1) \int_{m \in N} a_m \, d\mu(m).$$

For part 1, set $T^0 = 0$ and $T^+ = 2$. The equilibrium under T^0 is $a_n^*(T^0) = 1$ for all n with payoffs $-K$, while under T^+ it is $a_n^*(T^+) = 0$ for all n with payoffs 0. Thus $\delta_n(T^+, T^0) = K$. The direct effect from T^0 to T^+ is $v_n((\hat{a}_n, a^0_{-n}); T^+) - v_n(a^0; T^0) = -(K + 1) - (-K) = -1$, where $\hat{a}_n = 0$ and the integral remains 1 since a single agent has measure zero. The equilibrium effect is $K + 1$. For $\eta_n < \frac{1}{K+1}$ for all n, the perceived gain $\tilde{\delta}_n(T^+, T^0; \eta_n) = -1 + \eta_n(K + 1) < 0$. Since f satisfies consensus, T^+ is infeasible.

For part 2, set $T^0 = 2$ and $T^- = 0$. Symmetric arguments show that for $\eta_n < \frac{1}{K+1}$, the perceived gain is positive while the true gain is $-K$. Since f satisfies consensus, T^- is feasible. □

A.4 Proposition 3

Proof. WLOG let $p^0 = 0$, and let $p > 0$,[41] with associated Nash equilibrium $a^p = a^*(p)$ and $\alpha^t = \alpha(a^p)$ (notice that α^p is decreasing in p).

41. It is WLOG since we can always redefine $t' = t - t^0$.

Lemma 1 (Unbounded Action)**.** *For any tax rate $p > 0$, any aggregator α, and any action $a_n \in \{0, \ldots, M\}$, there exists a threshold $\overline{\theta}$ such that if $\theta_n > \overline{\theta}$, $a_n^*(\alpha, p) \geq a_n$.*

Proof. It suffices to show that for any $p > 0$ and α, some type chooses action M. An agent prefers M to $M-1$ if the marginal benefit exceeds p: $\nu(M, \alpha; \theta) - \nu(M-1, \alpha; \theta) > p$.

By the Assumption 3, this marginal benefit grows at least linearly with type:

$$[\nu(M, \alpha; \theta) - \nu(M-1, \alpha; \theta)] \geq [\nu(M, \alpha; 0) - \nu(M-1, \alpha; 0)] + \epsilon\theta.$$

As $\theta \to \infty$, the right-hand side grows without bound. Therefore, if θ_{max} is sufficiently large, there exists $\overline{\theta}$ such that for all $\theta > \overline{\theta}$, the marginal benefit exceeds p. For these types, the optimal action is M. □

Let $\hat{a}_n = \text{argmax}_{a_n \in A_n}\{w(a_n, \alpha^0; \theta_n) - pa_n\}$. Since $\alpha^p < M$, by Lemma 1, there exists $\overline{\theta'}$ such that if $\theta > \overline{\theta'}$, $\hat{a}_n > \alpha^p$.

The direct effect is given by

$$\nu(a_n^0, \alpha^0; \theta_n) - \nu(\hat{a}_n, \alpha^0; \theta_n) - p(\hat{a} - \alpha^p).$$

By best-response, the first part is weakly negative. The second part is strictly negative for $\hat{a}_n > \alpha^p$, i.e., if $\theta > \overline{\theta'}$. Therefore, for $\theta_n > \overline{\theta'}$, $\tilde{\delta}_n(T, T^0; 0) < 0$.

The actual gain for agent n is given by

$$\nu(a_n^p, \alpha^p; \theta_n) - \nu(a^0, \alpha^0; \theta_n) - p(a_n^p - \alpha^p)$$

We then have,

$$\begin{aligned}\nu(a_n^p, \alpha^p; \theta_n) - \nu(a^0, \alpha^0; \theta_n) - p(a_n^p - \alpha^p) &\geq \nu(a_n^0, \alpha^t; \theta_n) - \nu(a^0, \alpha^0; \theta_n) - p(a_n^0 - \alpha^p) \\ &\geq \nu(a_n^0, \alpha^p; \theta_n) - \nu(a^0, \alpha^0; \theta_n) - p(M - \alpha^p)\end{aligned}$$

Where the first inequality follows from best response, and the second from the definition of the action space. Using the same arguments as before, we can show that the RHS grows without bound, and in particular there exists $\overline{\theta''}$ such that for all $\theta \geq \overline{\theta''}$, the actual gain is positive i.e., $\delta(T, T^0) > 0$.

Therefore for all n with $\theta_n > \overline{\theta} \equiv \max\{\overline{\theta'}, \overline{\theta''}\}$, $\delta(T, T^0) > 0$ but $\tilde{\delta}_n(T, T^0; 0) < 0$.

To derive the bound on η, notice that

$$\tilde{\delta}_n(T, T^0; \eta) = \tilde{\delta}_n(T, T^0; 0) + \eta[\delta_n(T, T^0) - \tilde{\delta}_n(T, T^0; 0)]$$

therefore for $\eta < \frac{-\tilde{\delta}_n(T, T^0; 0)}{\delta_n(T, T^0) - \tilde{\delta}_n(T, T^0; 0)} \equiv \overline{\eta}(\theta_n)$ we have $\tilde{\delta}_n(T, T^0; \eta) < 0$, as needed. □

A.5 Result 2

Proof. From Result 1, the equilibrium cutoff θ^0 satisfies:

$$\theta^0 Q(1 - F(\theta^0)) = b$$

Clearly, any optimum must be cutoff based. Total welfare from an allocation where types $\theta > \theta'$ drive is given by:

$$W(\theta') = \int_{\theta'}^{\infty} [\theta Q(1 - F(\theta)) - b] dF(\theta)$$

The first-order condition yields:

$$\theta^* Q(1 - F(\theta^*); I) = b - Q'(1 - F(\theta^*)) \cdot \int_{\theta^*}^{\infty} \theta dF(\theta)$$

Since $Q' < 0$ and $\int_{\theta^*}^{\infty} \theta dF(\theta) > 0$:

$$\theta^* Q(1 - F(\theta^*)) > b = \theta^0 Q(1 - F(\theta^0))$$

Since $\theta \mapsto \theta Q(1 - F(\theta))$ is strictly increasing, we have $\theta^* > \theta^0$, which implies $\alpha^0 = 1 - F(\theta^0) > 1 - F(\theta^*) = \alpha^*$.

The gap $b - \theta^* Q(1 - F(\theta^*)) = Q'(1 - F(\theta^*)) \cdot \mathbb{E}[\theta | \theta \geq \theta^*] \cdot (1 - F(\theta^*))$ represents the marginal congestion externality, i.e., the social cost imposed on existing drivers by an additional driver. Setting the toll $p^* = -Q'(1 - F(\theta^*)) \cdot \mathbb{E}[\theta | \theta \geq \theta^*] \cdot (1 - F(\theta^*))$ internalizes this externality and achieves the social optimum. □

A.6 Theorem 1

Proof. Let $\mathcal{E}$ be a quasi-linear economy. Fix a status quo T^0 and a monotonic feasibility rule f. Let T^* be a policy that is politically feasible with respect to f under T^0 and induces an equilibrium $a^*(T^*) = a^*$. Denote the distinct status-quo equilibrium $a(T^0) = a^0$. Let

$$\tilde{T}_n(a) = T_n^*(a_n) + h_n(a)$$

where $h_n(a) = k \cdot \mathbf{1}\{a = a^0 \text{ a.e.}\}$.

Starting with outcome equivalence, since each agent is measure zero, by quasi-linearity and the definition of equilibrium we have:

$$\arg\max_{a_n \in A_n} w_n(a_n, a_{-n}^*; \tilde{T}) = \arg\max_{a_n \in A_n} w_n(a_n, a_{-n}^*; T^*) + h_n(a^*) = \arg\max_{a_n \in A_n} w_n(a_n, a^*; T^*) = a_n^*.$$

Therefore, if a^* is an equilibrium under T^*, it must also be an equilibrium under $\tilde{T}$.

To show deficit equivalence, since a^* and a^0 are distinct, no contingent compensation is triggered in equilibrium, i.e., $h_n(a_n^*, a_{-n}^*) = 0$ for all $n \in N$, which implies $\tilde{T}_n(a^*) = T_n^*(a^*)$ for all n, and thus $D_{T^*}^* = D_{\tilde{T}}^*$.

Finally, we need to show that $\tilde{T}$ is EN-robust. By monotonicity of f, it suffices to show $\tilde{\delta}_n(\tilde{T}, T^0; \eta) \geq \delta_n(\tilde{T}, T^0)$ for all $n \in N$ and all $\boldsymbol{\eta} \in [0,1]^N$.

For all n and all $\eta \in [0,1]$, we have

$$\tilde{\delta}_n(\tilde{T}, T^0; \eta) = \delta_n(T^*, T^0; \eta) + (1-\eta)k.$$

Therefore,

$$\begin{aligned}\tilde{\delta}_n(\tilde{T}, T^0; \eta) &= \eta\delta_n(\tilde{T}, T^0) + (1-\eta)[\tilde{\delta}_n(\tilde{T}, T^0; 0)] \\ &= \eta\delta_n(T^*, T^0) + (1-\eta)[\tilde{\delta}_n(T^*, T^0; 0) + k].\end{aligned}$$

Notice that since w_n is bounded, so is $\tilde{\delta}_n(T_n, T_n^0; \eta)$. Therefore, there exists a sufficiently large k that ensures

$$\tilde{\delta}_n(T^*, T^0; 0) + k \geq \delta_n(T^*, T^0),$$

which implies

$$\tilde{\delta}_n(\tilde{T}_n, T_n^0; \eta) \geq \delta_n(\tilde{T}_n, T_n^0)$$

for all $n \in N$ and all $\boldsymbol{\eta} \in [0,1]^N$, as needed.

Finite Economies. To extend the proof for finite economies, let N be a finite set. In finite economies, profiles a and a' are said to be distinct when they differ by at least two actions, i.e., when $|a_n \neq a_n'| > 1$.

Let $h_n(a) = k \cdot \mathbf{1}\{a_{-n} = a_{-n}^0\}$, i.e., a compensation triggered when others' status-quo behavior persists. Notice that h_n depends on others' actions, and therefore does not affect best response. In addition, since a^* and a^0 are distinct, $a_{-n}^* \neq a_{-n}^0$ for all n, which implies that no payment is triggered under the equilibrium profile a^*.

□

A.7 Theorem 2

Proof. Following the same steps as in the proof of Theorem 1, let

$$h_n(a) = k(a_n^0) \cdot \mathbf{1}\{a = a^0 \text{ a.e.}\},$$

where $k(a_n^0) = T_n^*(a_n^0) - T_n^0(a^0)$. Note that relative to the EN-robust policy from Theorem 1, constructing $\tilde{T}$ requires only one additional piece of information: the difference between

what agent n pays under the status quo versus under T^*, holding their behavior fixed.

Treating agents' status behavior a_n^0 as a fixed parameter, outcome and (on-path) deficit equivalence follow from the same steps as in the proof of Theorem 1. To establish EN-robustness with respect to the weighted-support based rule f, we must show that if $\delta_n(T, T^0) \geq 0$, then $\tilde{\delta}_n(\tilde{T}, T^0; \eta) \geq 0$ for all $\eta \in [0, 1]$.

Suppose $\delta_n(T^*, T^0) \geq 0$. From the proof of Theorem 1, we have

$$\begin{aligned}\tilde{\delta}_n(\tilde{T}, T^0; \eta) &= \eta\delta_n(T^*, T^0) + (1-\eta)[\tilde{\delta}_n(T^*, T^0; 0) + k(a_n^0)] \\ &= \eta\delta_n(T^*, T^0) + (1-\eta)[\tilde{\delta}_n(T^*, T^0; 0) + T_n^*(a_n^0) - T_n^0(a^0)]\end{aligned}$$

and also

$$\tilde{\delta}_n(T^*, T^0; 0) \geq T_n^*(a_n^0) + T_n^0(a^0).$$

Therefore, if $\delta_n(T^*, T^0) \geq 0$, then $\tilde{\delta}_n(\tilde{T}, T^0; \eta) \geq 0$ for all $\eta \in [0, 1]$, as required.

Finally, to characterize the deficit: for any $a \neq a^0$ almost surely, we have $D_{\tilde{T}}(a) = D_{T^*}(a)$ by construction. For a^0, we have

$$D_{\tilde{T}}(a^0) = \int_{n \in N} [T_n^*(a_n^0) + k_n] d\mu(n) = \int_{n \in N} T_n^0(a^0) d\mu(n) = D_{T^0}(a^0)$$

as needed.

Finite Economies. Applying the same modification to h_n as in the proof of Theorem 1 maintains the same properties. □

B Aggregative Economies with Uncertainty

This section extends the model to allow for uncertainty. I focus on aggregative economies. There is a random state $\omega \in \Omega$ drawn from distribution H. The state captures common shocks to preferences, technology, or the environment. Agent n's state-contingent utility is

$$\nu_n(a_n, \alpha; \omega) + T_n(a_n, \alpha),$$

where $\nu_n : A_n \times \mathcal{Q} \times \Omega \to \mathbb{R}$ and transfers are state-independent. To simplify notation, multilateral policies are further restricted to depend only on the agent's action and the aggregate, not the full action profile.

The timing is as follows: the planner first commits to policy $T \in \mathbb{T}$; a state $\omega \in \Omega$ is realized and commonly observed; agents simultaneously choose actions. An *aggregative economy with uncertainty* $\hat{\mathcal{E}} = \langle N, A, \nu, \mathbb{T}, \alpha, \Omega, H \rangle$ specifies a family of policy-parameterized state-contingent aggregative games, where policy T and state ω jointly determine payoffs.

Under standard regularity conditions, each game in the family has a Nash equilibrium in pure or mixed strategies. Let $\Pi_T \in \Delta(\Omega \times \mathcal{Q})$ denote the joint distribution of (ω, α) under policy T, where $\omega \sim H$ and α is drawn from the equilibrium distribution conditional on ω.[42]

Taking aggregate α as given, agent n's best response to policy T in state ω is

$$a_n^*(\alpha, T_n, \omega) = \arg\max_{a_n \in A_n} \{\nu_n(a_n, \alpha; \omega) + T_n(a_n, \alpha)\}.$$

The *indirect utility* is

$$\nu_n^*(\alpha, T_n, \omega) = \nu_n(a_n^*(\alpha, T_n, \omega), \alpha; \omega) + T_n(a_n^*(\alpha, T_n, \omega), \alpha).$$

Agent n's expected utility under policy T is

$$V_n(T) = \mathbb{E}_{\Pi_T}\left[\nu_n^*(\alpha, T_n, \omega)\right],$$

and the *expected gain* from policy T over status quo T^0 is

$$\delta_n(T, T^0) = V_n(T) - V_n(T^0).$$

B.1 Equilibrium Neglect

Equilibrium neglect extends to this setting as under-appreciation of the policy-induced shift in the aggregate distribution. When evaluating policy T against status quo T^0, the agent perceives the joint distribution as

$$\tilde{\Pi}_{T,T^0}(\cdot\,; \eta) = \eta\, \Pi_T(\cdot) + (1 - \eta)\, \Pi_{T^0}(\cdot).$$

Agents correctly perceive the exogenous state distribution H but distort the aggregate distribution α toward the status quo. The *perceived* expected utility under policy T is

$$\tilde{V}_n(T, T^0; \eta) = \mathbb{E}_{\tilde{\Pi}_{T,T^0}}\left[\nu_n^*(\alpha, T_n, \omega)\right],$$

42. If the game has more than one equilibrium, Π may incorporate any arbitrary selection criterion which can include mixture of equilibria. Notice that unlike in the main text, the existence of a pure Nash equilibrium is not assumed.

and the *perceived* expected gain is $\tilde{\delta}_n(T, T^0; \eta) = \tilde{V}_n(T, T^0; \eta) - V_n(T^0)$, which can be decomposed as:

$$\tilde{\delta}_n(T, T^0; \eta) = \underbrace{\mathbb{E}_{\Pi_{T^0}}\left[\nu_n^*(\alpha, T_n, \omega) - \nu_n^*(\alpha, T_n^0, \omega)\right]}_{\text{direct effect}} + \eta \underbrace{\left(\mathbb{E}_{\Pi_T}\left[\nu_n^*(\alpha, T_n, \omega)\right] - \mathbb{E}_{\Pi_{T^0}}\left[\nu_n^*(\alpha, T_n, \omega)\right]\right)}_{\text{equilibrium effect}}.$$

As in the baseline version, the direct effect captures the welfare change from adopting T while holding others' behavior, and hence the distribution of the equilibrium aggregate, fixed at Π_{T^0}. The equilibrium effect captures the shift in the distribution of the aggregate from Π_{T^0} to Π_T.

The equilibrium neglect parameter $\eta \in [0, 1]$ may represent an "awareness probability" of considering equilibrium adjustments; attenuated perception of behavioral responses due to complexity or salience. It may also represent a general underestimation of the population elasticity with respect to policy change via an attenuated perceived divergence between Π_{T^0} and Π_T.

B.2 EN-robust policies

We now extend the EN-robustness results to allow for uncertainty. The key idea is to construct contingent compensations whose expected value offsets the gap between perceived and actual benefits, and whose expected deficit under the new equilibrium is zero. To achieve this, contingent compensations scale linearly with the distance from the expected equilibrium aggregate.

The objectives from the baseline model extend naturally. Two policies T and T' are *distributionally outcome equivalent* if $\Pi_T = \Pi_{T'}$, i.e., if they induce the same distribution over the aggregate. The *expected equilibrium deficit* of T is

$$\hat{D}_T^* = \int_{n \in N} \mathbb{E}_{\Pi_T}\left[T_n(a_n^*(\alpha, T_n, \omega), \alpha)\right] d\mu(n),$$

where the expectation is over $(\omega, \alpha) \sim \Pi_T$. Two policies are *deficit equivalent in expectation* if they induce the same expected equilibrium deficit.

Theorem 3. *Let $\hat{\mathcal{E}}$ be an aggregative economy with uncertainty, T^0 a status quo policy, f a monotonic feasibility rule, and T^* an anonymous unilateral policy. Suppose $\Pi_{T^*} \neq \Pi_{T^0}$. If T^* is politically feasible under T^0 with respect to f when $\boldsymbol{\eta} = \mathbf{1}$, then there exists a multilateral policy $\tilde{T}$ that is distributionally outcome equivalent and deficit equivalent in expectation to T^* and is politically feasible with respect to f for all $\boldsymbol{\eta} \in [0, 1]^N$.*

B.2.1 Off-path Expected Deficit

For the off-path deficit result, compensations must depend on past actions. In the stochastic setting, they must depend on the *distribution* of past actions under the status quo.

Formally, a *history-based multilateral policy* is a function $T_n : \Delta(A_n) \times A_n \times \mathcal{Q} \to \mathbb{R}$ that depends on the agent's status quo action distribution $\pi_n^0 \in \Delta(A_n)$, current action, and aggregate.

To allow for off-path deficit, the expected deficit of T at Π is

$$\hat{D}_T(\Pi) = \int_{n \in N} \mathbb{E}_\Pi \left[T_n(\pi_n^0, a_n^*, \alpha) \right] d\mu(n),$$

where a_n^* and α denote the equilibrium action and aggregate, and the expectation is over $(\omega, \alpha) \sim \Pi$.

Theorem 4. *Let $\hat{\mathcal{E}}$ be an aggregative economy with uncertainty, T^0 a status quo policy, f a weighted-support-based rule, and T^* an anonymous unilateral policy. Suppose $\Pi_{T^*} \neq \Pi_{T^0}$. If T^* is politically feasible under T^0 with respect to f when $\boldsymbol{\eta} = \mathbf{1}$, then there exists a history-based multilateral policy $\tilde{T}$ that is outcome and deficit equivalent in expectation to T^* and is politically feasible with respect to f for all $\boldsymbol{\eta} \in [0,1]^N$. Furthermore, $\tilde{T}$ may be constructed so that $\hat{D}_{\tilde{T}}(\Pi_{T^0}) = \hat{D}^*_{T^0}$.*

B.3 Proofs

B.3.1 Theorem 3

Proof. Let $\hat{\mathcal{E}}$ be an aggregative economy with uncertainty. Fix status quo T^0, monotonic feasibility rule f, and unilateral policy T^* that is politically feasible with respect to f under T^0. Let Π_{T^*} denote the equilibrium distribution under T^* with expected aggregate $\alpha^* = \mathbb{E}_{\Pi_{T^*}}[\alpha]$, and let Π_{T^0} denote the status quo equilibrium distribution with expected aggregate $\alpha^0 = \mathbb{E}_{\Pi_{T^0}}[\alpha] \neq \alpha^*$.

Define the multilateral policy

$$\tilde{T}_n(a_n, \alpha) = T_n^*(a_n) + k \cdot \frac{\alpha - \alpha^*}{\alpha^0 - \alpha^*},$$

where $k > 0$ is to be determined.

Distributional outcome equivalence: By quasi-linearity, adding a function of α alone does not affect best responses when agents take α as given. Thus for any $\omega \in \Omega$ and $\alpha \in \mathcal{A}$,

$$a_n^*(\alpha, \tilde{T}_n, \omega) = a_n^*(\alpha, T_n^*, \omega),$$

which implies $\Pi_{\tilde{T}} = \Pi_{T^*}$.

Deficit equivalence in expectation: Outcome equivalence implies

$$\begin{aligned}\hat{D}^*_{\tilde{T}} &= \int_{n\in N} \mathbb{E}_{\Pi_{\tilde{T}}}\left[T^*_n(a^*_n) + k\frac{\alpha-\alpha^*}{\alpha^0-\alpha^*}\right] d\mu(n) \\ &= \int_{n\in N} \mathbb{E}_{\Pi_{T^*}}\left[T^*_n(a^*_n)\right] d\mu(n) + k\underbrace{\mathbb{E}_{\Pi_{T^*}}\left[\frac{\alpha-\alpha^*}{\alpha^0-\alpha^*}\right]}_{=0} \\ &= \hat{D}^*_{T^*}.\end{aligned}$$

EN-robustness: By monotonicity of f, it suffices to show $\tilde{\delta}_n(\tilde{T}, T^0; \eta) \geq \delta_n(\tilde{T}, T^0)$ for all $n \in N$ and $\eta \in [0,1]$.

Since $u_n(\alpha, \tilde{T}_n, \omega) = u_n(\alpha, T^*_n, \omega) + k\frac{\alpha-\alpha^*}{\alpha^0-\alpha^*}$, we have

$$\begin{aligned}\tilde{\delta}_n(\tilde{T}, T^0; \eta) &= \eta\mathbb{E}_{\Pi_{\tilde{T}}}[u_n(\alpha, \tilde{T}_n, \omega)] + (1-\eta)\mathbb{E}_{\Pi_{T^0}}[u_n(\alpha, \tilde{T}_n, \omega)] - V_n(T^0) \\ &= \eta\mathbb{E}_{\Pi_{T^*}}[u_n(\alpha, T^*_n, \omega)] + (1-\eta)\mathbb{E}_{\Pi_{T^0}}[u_n(\alpha, T^*_n, \omega)] + (1-\eta)k - V_n(T^0) \\ &= \eta\delta_n(T^*, T^0) + (1-\eta)[\tilde{\delta}_n(T^*, T^0; 0) + k],\end{aligned}$$

where the second equality uses $\mathbb{E}_{\Pi_{T^*}}\left[\frac{\alpha-\alpha^*}{\alpha^0-\alpha^*}\right] = 0$ and $\mathbb{E}_{\Pi_{T^0}}\left[\frac{\alpha-\alpha^*}{\alpha^0-\alpha^*}\right] = 1$, and the third follows from the definitions of $\delta_n(T^*, T^0)$ and $\tilde{\delta}_n(T^*, T^0; 0)$.

Since all relevant functions are bounded, there exists k sufficiently large such that

$$\tilde{\delta}_n(\tilde{T}, T^0; \eta) \geq \delta_n(\tilde{T}, T^0)$$

for all $n \in N$ and $\eta \in [0,1]$, as required. □

B.3.2 Theorem 4

Proof. Let $\pi^0_n \in \Delta(A_n)$ denote the marginal distribution of agent n's equilibrium action under the status quo, induced by Π_{T^0}. Define the history-based multilateral policy

$$\tilde{T}_n(\pi^0_n, a_n, \alpha) = T^*_n(a_n) + h_n(\alpha, \pi^0_n),$$

where

$$h_n(\alpha, \pi^0_n) = \mathbb{E}_{\pi^0_n}[T^0_n(a_n) - T^*_n(a_n)] \cdot \frac{\alpha-\alpha^*}{\alpha^0-\alpha^*}.$$

Thus we replace the constant k from Theorem 3 with a history-dependent compensation equal to the expected difference in transfers under the status quo action distribution.

Outcome and deficit equivalence: These follow directly from the construction, as in Theorem 3.

EN-robustness: We need to show that $\delta_n(T^*, T^0) \geq 0$ implies $\tilde{\delta}_n(\tilde{T}, T^0; \eta) \geq 0$ for all $\eta \in [0, 1]$. As before,

$$\tilde{\delta}_n(\tilde{T}, T^0; \eta) = \eta \delta_n(T^*, T^0) + (1 - \eta) \left[\tilde{\delta}_n(T^*, T^0; 0) + \mathbb{E}_{\pi_n^0}[T_n^0(a_n) - T_n^*(a_n)] \right].$$

The first term is weakly positive by assumption. For the second term, observe that

$$\begin{aligned}
\tilde{\delta}_n(T^*, T^0; 0) &= \mathbb{E}_{\Pi_{T^0}}[w_n(a_n^*(T_n^*), \alpha, \omega) + T_n^*(a_n^*(T_n^*))] \\
&\quad - \mathbb{E}_{\Pi_{T^0}}[w_n(a_n^*(T_n^0), \alpha, \omega) + T_n^0(a_n^*(T_n^0))] \\
&\geq \mathbb{E}_{\Pi_{T^0}}[w_n(a_n^*(T_n^0), \alpha, \omega) + T_n^*(a_n^*(T_n^0))] \\
&\quad - \mathbb{E}_{\Pi_{T^0}}[w_n(a_n^*(T_n^0), \alpha, \omega) + T_n^0(a_n^*(T_n^0))] \\
&= \mathbb{E}_{\Pi_{T^0}}[T_n^*(a_n^*(T_n^0)) - T_n^0(a_n^*(T_n^0))] \\
&= \mathbb{E}_{\pi_n^0}[T_n^*(a_n) - T_n^0(a_n)],
\end{aligned}$$

where the inequality follows from best response and the last equality uses the fact that $a_n^*(T_n^0)$ has marginal distribution π_n^0 under Π_{T^0}. Therefore, $\tilde{\delta}_n(T^*, T^0; 0) + \mathbb{E}_{\pi_n^0}[T_n^0(a_n) - T_n^*(a_n)] \geq 0$, which implies $\tilde{\delta}_n(\tilde{T}, T^0; \eta) \geq 0$ for all $\eta \in [0, 1]$.

Off-path deficit: The expected deficit of $\tilde{T}$ under Π_{T^0} is

$$\begin{aligned}
\hat{D}_{\tilde{T}}(\Pi_{T^0}) &= \int_{n \in N} \mathbb{E}_{\Pi_{T^0}}[\tilde{T}_n(\pi_n^0, a_n^*, \alpha)] d\mu(n) \\
&= \int_{n \in N} \mathbb{E}_{\Pi_{T^0}}[T_n^*(a_n^*)] d\mu(n) + \int_{n \in N} \mathbb{E}_{\pi_n^0}[T_n^0(a_n) - T_n^*(a_n)] \cdot \underbrace{\mathbb{E}_{\Pi_{T^0}}\left[\frac{\alpha - \alpha^*}{\alpha^0 - \alpha^*} \right]}_{=1} d\mu(n) \\
&= \int_{n \in N} \mathbb{E}_{\pi_n^0}[T_n^*(a_n)] d\mu(n) + \int_N \mathbb{E}_{\pi_n^0}[T_n^0(a_n) - T_n^*(a_n)] d\mu(n) \\
&= \int_{n \in N} \mathbb{E}_{\pi_n^0}[T_n^0(a_n)] d\mu(n) \\
&= \int_{n \in N} \mathbb{E}_{\Pi_{T^0}}[T_n^0(a_n^*)] d\mu(n) = \hat{D}_{T^0}(\Pi_{T^0}) = \hat{D}_{T^0}^*,
\end{aligned}$$

as required. □

Table C.1: Data Sources for Support Polls

Location	Period (Date)	Source	Sampling	N
Bergen	Before (December 1985)	Larsen (1995)	–	–
	After (December 1986)	Larsen (1995)	–	–
Oslo	Before (1989)	Odeck and Bråthen (1997)	Phone	1,100
	After (1991)	Odeck and Bråthen (1997)	Phone	1,100
Trondheim	Before (April 1991)	Tretvik (2003)	–	–
	After (December 1991)	Tretvik (2003)	–	–
London	Before (February 2003)	Transport for London (2004)	Phone	1,000
	After (March 2003)	Transport for London (2004)	Phone	1,000
Stockholm	Before (2005)	Eliasson (2014)	Phone	1,600
	After (2006)	Eliasson (2014)	Phone	1,600
Gothenburg	Before (2012)	Börjesson et al. (2016)	Mail	1,500
	After (2013)	Börjesson et al. (2016)	Mail	1,500
New York	Before (December 2024)	Siena College Research Institute (2024)	Phone/online	834
	After (March 2025)	Siena College Research Institute (2025)	Phone/online	806

Note: This table lists the data sources for the public support polls presented in Figure 2. "Before" and "After" refer to polling conducted before and after the implementation of congestion pricing in each location.

C Existing Polls: Further Details

C.1 Data Sources for Figure 2

C.2 Siena Polls: Details and Additional Results Additional Results

Table C.2: Siena Polls: Details

Poll Name	Date	Respondents	Question Wording
SNY 0424	April 15–17, 2024	806 NYS Reg. Voters	*Do you support or oppose the MTA congestion pricing plan—expected to be implemented soon—that will charge a $15 per car toll on most passenger vehicles that enter Manhattan south of 60th Street, in order to fund mass transit improvements?*
SNY 1224	Dec 2–5, 2024	834 NYS Reg. Voters	*In June, Governor Hochul put a $15 congestion pricing toll plan for Manhattan on hold. Last month, after the election, Governor Hochul reinstituted a $9 congestion pricing plan, set to begin in January. Do you support or oppose the Governor´s decision to implement the $9 congestion pricing toll for Manhattan?*
SNY 0325	Mar 2–6, 2025	806 NYS Reg. Voters	*(Statements rotated) Governor Hochul says that congestion pricing in Manhattan is working, reducing traffic and raising revenue for the MTA, and it should remain. President Trump says that congestion pricing is nothing more than an unfair tax on working commuters, and it should be eliminated. Do you think congestion pricing: Should remain/Should be eliminated/In the middle*
ICS 0325	March 17–24	1,005 NYS Residents	*Please indicate whether you support or oppose the following proposals and initiatives being discussed by state policy makers and legislators. New York State implementing policies that aim to reduce vehicle use, such as congestion pricing or restricted driving zones.*
SNY 0525	May 12–15, 2025	805 NYS Reg. Voters	*Same as SNY 0325*

Note: Details on the five polls in the dataset. Figure 3 includes SNY1224 and SNY0325, which were conducted a month before and two month after the implementation of the policy.

Table C.3: Siena Polls: Congestion Pricing Support Increase by Manhattan Commute

	SNY Dec 24–Mar 25			All Polls		
	(1)	(2)	(3)	(4)	(5)	(6)
After	0.07	0.09	0.11	0.17	0.16	0.27
	(0.06)	(0.06)	(0.91)	(0.04)	(0.04)	(0.54)
High Commute Intensity	−0.07	−0.15	−0.12	−0.04	−0.08	−0.07
	(0.06)	(0.06)	(0.07)	(0.04)	(0.04)	(0.05)
Republican		−0.55	−0.48		−0.50	−0.41
		(0.05)	(0.07)		(0.03)	(0.05)
After × High Commute	0.25	0.23	0.19	0.23	0.19	0.17
	(0.09)	(0.08)	(0.09)	(0.06)	(0.05)	(0.06)
After × Republican			−0.16			−0.17
			(0.10)			(0.07)
Demographic Controls		✓	✓		✓	✓
Fully Interacted			✓			✓
Observations	1,640	1,640	1,640	3,869	3,869	3,869
R^2	0.02	0.16	0.18	0.03	0.15	0.16

Note: SNY Dec 2024–Mar 2025 sample includes SNY1224 and SNY0325, which were conducted a month before and two month after the implementation of the policy, with the same sampling criterion. All polls sample includes all five available waves (SNY0424, SNY1224, SNY0325, ICS0325, SNY0525). Further details are in Table C.2. High commute intensity defined as above-median share of county workers commuting to Manhattan. Demographic controls include income, employment, homeownership, gender, age, race, education, and religion. Unweighted regressions. Standard errors in parentheses.

Table C.4: Siena Polls: Congestion Pricing Support Increase by Manhattan Commute (Weighted)

	SNY Dec 24– Mar 25			All Polls		
	(1)	(2)	(3)	(4)	(5)	(6)
After	−0.01	0.004	−0.11	0.16	0.15	0.08
	(0.06)	(0.06)	(0.98)	(0.04)	(0.04)	(0.56)
High Commute Intensity	−0.08	−0.20	−0.17	−0.02	−0.09	−0.07
	(0.06)	(0.06)	(0.07)	(0.04)	(0.04)	(0.05)
Republican		−0.44	−0.35		−0.40	−0.31
		(0.05)	(0.07)		(0.04)	(0.05)
After × High Commute	0.30	0.29	0.26	0.20	0.18	0.15
	(0.08)	(0.08)	(0.10)	(0.06)	(0.06)	(0.06)
After × Republican			−0.20			−0.16
			(0.10)			(0.07)
Demographic Controls		✓	✓		✓	✓
Fully Interacted			✓			✓
Observations	1,640	1,640	1,640	3,869	3,869	3,869
R^2	0.02	0.11	0.14	0.03	0.11	0.12

Note: Dec 2024–Mar 2025 sample includes SNY1224 and SNY0325. All polls sample includes all five available waves (SNY0424, SNY1224, SNY0325, ICS0325, SNY0525). High commute intensity defined as above-median share of county workers commuting to Manhattan. Demographic controls include partisanship, income, employment, homeownership, gender, age, race, education, and religion. Regressions weighted by survey sampling weights. Standard errors in parentheses.

D Survey Instruments

INTRODUCTION

Thank you for agreeing to participate in our new AmeriSpeak survey!

This survey is about economic policy.

As always, your answers are confidential.

Baseline Support

Some cities are considering implementing “congestion pricing," which charges drivers a fee to enter busy downtown areas during peak hours.

- **Supporters** cite research from other cities suggesting it can cut traffic by about 15–20% and reduce travel times by 10–15 minutes.
- **Opponents** argue it won't meaningfully reduce congestion and will unfairly burden workers who must drive.

Consider the following policy for [CITY].

- **Toll:** Drivers entering downtown [CITY] during peak hours would have to pay a **$15** toll.
- **Revenue:** The revenue would fund transportation improvements.

How much would you support or oppose this policy in [CITY]?

01 Strongly support
02 Somewhat support
03 Neither support nor oppose
04 Somewhat oppose
05 Strongly oppose

Blocks 1 and 2 are order-randomized.

Block 1: Perceived Effectiveness + Confidence

[50% of respondents see this question:]
If congestion pricing were implemented according to the proposal you just saw, how do you think it would affect driving time from your home to downtown [CITY]?

(If you live in downtown [CITY], please answer how it would affect driving time for your typical trips within downtown.)

01 Driving time would increase by more than 15 minutes
02 Driving time would increase by 11–15 minutes
03 Driving time would increase by 6–10 minutes
04 Driving time would increase by 1–5 minutes
05 Driving time would not change
06 Driving time would decrease by 1–5 minutes
07 Driving time would decrease by 6–10 minutes
08 Driving time would decrease by 11–15 minutes
09 Driving time would decrease by more than 15 minutes

[[50% of respondents see this question:]
In 2024, the typical weekday evening commute around 5:00 PM through the Holland Tunnel—connecting New Jersey to Lower Manhattan—took an average of 23 minutes (compared to about 8 minutes with no traffic).
On January 5, 2025, a **$9** daytime toll was introduced for cars entering Lower Manhattan.
During the first two months *following this change* (i.e., while the **$9** toll was in place), what do you estimate was the typical weekday commute time at 5:00 PM through the Holland Tunnel?

01 36 minutes or more
02 31-35 minutes
03 26-30 minutes
04 21-25 minutes
05 16-20 minutes
06 11-15 minutes
07 10 minutes or less

[Own-commute cognitive uncertainty:]
You indicated that if congestion pricing were implemented according to the proposal you just saw, you think driving time from your home to downtown [CITY] would **[ANSWER]**.
How certain are you that driving time would **[ANSWER]**?]

[Hollan tunnel cognitive uncertainty:]
You indicated that you estimate the typical weekday commute time at 5:00 PM through the Holland Tunnel during the two months following the introduction of the $9 toll to be **[ANSWER]**.
How certain are you that the typical weekday commute time for that period was **[ANSWER]**?

01 0% (very uncertain)
02 10%
03 20%

04 30%
05 40%
06 50%
07 60%
08 70%
09 80%
10 90%
11 100% (completely certain)

Block 2: Habits + Location

DRIVING.
How often do you drive into downtown [CITY]?
(If you live in downtown [CITY], please answer how often you drive out of the downtown area.)

01 Never
02 Less than once a month
03 1-3 times per month
04 1-2 times per week
05 3-4 times per week
06 5 or more times per week
07 Not applicable (I don't own or have access to a car)

WORKPLACE.
Which best describes your current workplace location?

01 Downtown [CITY]
02 Other part of [CITY]
03 Suburban office park
04 Small town/rural area
05 I work from home
06 I am not currently employed

2nd-stage support

[EN-robust Condition]

We would now like you to consider a **different** version of the congestion pricing policy. Any changes from the previous version will be larger in size and highlighted. Please review the new proposal carefully.

Some cities are considering implementing “congestion pricing," which charges drivers a fee to enter busy downtown areas during peak hours.

- **Supporters** cite research from other cities suggesting it can cut traffic by about 15–20% and reduce travel times by 10–15 minutes.
- **Opponents** argue it won't meaningfully reduce congestion and will unfairly burden workers who must drive.

Consider the following policy for [CITY].

- **Toll:** Drivers entering downtown [CITY] during peak hours would have to pay a **$15** toll.
- **Revenue:** The revenue would fund transportation improvements.
- **Monitoring and Assurance:** Traffic levels would be monitored and published monthly. If monthly traffic decreases by less than 8% (about half the typical 15–20% reduction) all workers who commute to work in the downtown area would receive a **$120** tax credit for that month, regardless of how they commute.
    - This is intended to ensure affected workers either benefit from shorter commute times or receive compensation if traffic does not decline as expected.

How much would you support or oppose this policy in [CITY]?

01 Strongly support
02 Somewhat support
03 Neither support nor oppose
04 Somewhat oppose
05 Strongly oppose

[Control Condition]

We would now like you to consider a **different** version of the congestion pricing policy. Any changes from the previous version will be larger in size and highlighted. Please review the new proposal carefully.

Some cities are considering implementing “congestion pricing," which charges drivers a fee to enter busy downtown areas during peak hours.

- **Supporters** cite research from other cities suggesting it can cut traffic by about 15–20% and reduce travel times by 10–15 minutes.

- **Opponents** argue it won't meaningfully reduce congestion and will unfairly burden workers who must drive.

Consider the following policy for [CITY].

- **Toll:** Drivers entering downtown [CITY] during peak hours would have to pay a **$15** toll.
- **Revenue:** The revenue would fund transportation improvements.
- **Monitoring:** Traffic levels would be monitored and published monthly.
    - This is intended to provide data about the program's effects on traffic patterns.

How much would you support or oppose this policy in [CITY]?

01 Strongly support
02 Somewhat support
03 Neither support nor oppose
04 Somewhat oppose
05 Strongly oppose

[Toll-reduction condition]

We would now like you to consider a **different** version of the congestion pricing policy. Any changes from the previous version will be larger in size and highlighted. Please review the new proposal carefully.

Some cities are considering implementing “congestion pricing," which charges drivers a fee to enter busy downtown areas during peak hours.

- **Supporters** cite research from other cities suggesting it can cut traffic by about 15–20% and reduce travel times by 10–15 minutes.
- **Opponents** argue it won't meaningfully reduce congestion and will unfairly burden workers who must drive.

Consider the following policy for [CITY].

- **Toll:** Drivers entering downtown **[CITY]** during peak hours would have to pay a ~~**$15**~~ **$9** toll.
- **Revenue:** The revenue would fund transportation improvements.

How much would you support or oppose this policy in [CITY]?

01 Strongly support
02 Somewhat support
03 Neither support nor oppose

04 Somewhat oppose
05 Strongly oppose

2nd-stage Belief

[If asked about Holland tunnel in 1st stage:]
If congestion pricing were implemented according to the proposal you just saw, how do you think it would affect driving time from your home to downtown [CITY]?
(If you live in downtown [CITY], please answer how it would affect driving time for your typical trips within downtown.)

01 Driving time would increase by more than 15 minutes
02 Driving time would increase by 11–15 minutes
03 Driving time would increase by 6–10 minutes
04 Driving time would increase by 1–5 minutes
05 Driving time would not change
06 Driving time would decrease by 1–5 minutes
07 Driving time would decrease by 6–10 minutes
08 Driving time would decrease by 11–15 minutes
09 Driving time would decrease by more than 15 minutes

[If asked about own commute in the first stage:]
In 2024, the typical weekday evening commute around 5:00 PM through the Holland Tunnel—connecting New Jersey to Lower Manhattan—took an average of 23 minutes (compared to about 8 minutes with no traffic).
On January 5, 2025, a **$9** daytime toll was introduced for cars entering Lower Manhattan.
During the first two months *following this change* (i.e., while the **$9** toll was in place), what do you estimate was the typical weekday commute time at 5:00 PM through the Holland Tunnel?

01 36 minutes or more
02 31-35 minutes
03 26-30 minutes
04 21-25 minutes
05 16-20 minutes
06 11-15 minutes
07 10 minutes or less

E Calculation of Budget-balanced Compensation

This appendix uses data from NYC to arrive at \$120 as a monthly compensation that maintains budget balance when congestion remains at its status-quo level. The two main assumptions required to make it applicable to other cities is that reduction in traffic volume would be similar in other cities, and that the ratio between workers and overall commuters is similar.

Revenue Calculation

For our calculations, we use the following notation:

- V: number of cars entering the city center on a typical weekday in the untolled scenario
- w: weekend traffic discount factor
- T': base toll (passenger vehicles)
- T: average toll (across vehicle types)
- $\tau(T)$: Traffic adjustment given toll
- r: average daily revenue
- $R = 365r$: annual revenue
- a: administrative costs and exemptions as a share of revenue

Given a toll T, the new level of traffic is $\tau(T) \cdot V$. Taking into account around 30% of days in a given year being holidays and weekends, the average daily discount factor (i.e., averaging across weekends and weekdays) is given by:

$$w' = 0.7 + 0.3w$$

The average daily revenue is then given by:

$$r = a \cdot T \cdot \tau(T) \cdot w' \cdot V$$

We assume a holiday/weekend discount factor $w = 0.7$, which implies $w' = 0.91$. For NYC, we have $V = 700,000$ (Metropolitan Transportation Authority (MTA) 2024a). The share of taxis entering Manhattan is approximately 0.35, and taxis pay on average

$\frac{1}{8}$ of the toll,[43] which implies:

$$T = 0.65T' + 0.35 \cdot \frac{1}{8} \cdot T' \approx 0.69T'$$

Solving for a:

$$a = \frac{R}{T' \cdot \tau(T) \cdot 441,919}$$

Imputed Deadweight Parameters

The table below imputes the deadweight parameter for different proposed tolling scenarios from the MTA's Environmental Assessment (Metropolitan Transportation Authority (MTA) 2024a; 2024b):

MTA EA Tolling Scenario	Base Toll (T')	MTA Projected Traffic ($\tau(T)$)	MTA Projected Revenue (R in billions)	Imputed Deadweight Parameter (a)
Adopted (original)	\$15	0.827	0.900	0.450
Adopted (revise phase 1)	\$9	0.876	0.500	0.393
Adopted (revise phase 2)	\$12	0.845	0.700	0.428
Adopted (revise phase 3)	\$15	0.827	0.900	0.450
A	\$9	0.846	1.060	0.863
B	\$10	0.843	0.830	0.610
C	\$14	0.827	1.100	0.589
D	\$19	0.813	1.340	0.538
E	\$23	0.801	1.480	0.498
F	\$23	0.817	1.020	0.337
G	\$12	0.840	1.100	0.677

Note: Imputed Deadweight Parameters for MTA Tolling Scenarios

Off-Path Compensation

Denote the number of workers who live outside the tolled area but are employed inside of it by E. The upper bound on total compensation per worker per day is therefore given by:

$$c = \tau(T) \cdot r \cdot E = T \cdot a \cdot w' \cdot V \cdot E$$

Denote the monthly compensation by $C = 31c$. There are about 1 million employees who commute to work in the Manhattan CBD,[44] which implies $VE = 0.7$. Plugging in our estimate from before and assuming $a = 0.5$ (approximately the average across all tolling scenarios), we get:

$$C(T) = 10T$$

For example, $C(15) = 150$. However, one may want to set aside a budget dedicated to public transit even off the equilibrium path. Setting this amount to $\frac{1}{5}$ of the revenue

43. In NYC, taxis pay $\frac{T}{12}$, and For-Hire Vehicles pay $\frac{T}{6}$, and I assume equal shares.

44. This article calculates this number from ACS data: https://rpa.org/appendix-cc, see the table titled "Number and Wages of Commuters Working in New York City." A similar number also appears in official documents: https://www.nyc.gov/assets/planning/download/pdf/planning-level/housing-economy/nyc-ins-and-out-of-commuting.pdf, see under "Regional Exchange of Workers by Subregion."

results in:

$$C(15) = 120.$$

F Survey: Additional Results

F.1 Details on Estimated Time Savings from NYC

Table F.1: Congestion Reduction Estimates by Manhattan Entrance Route

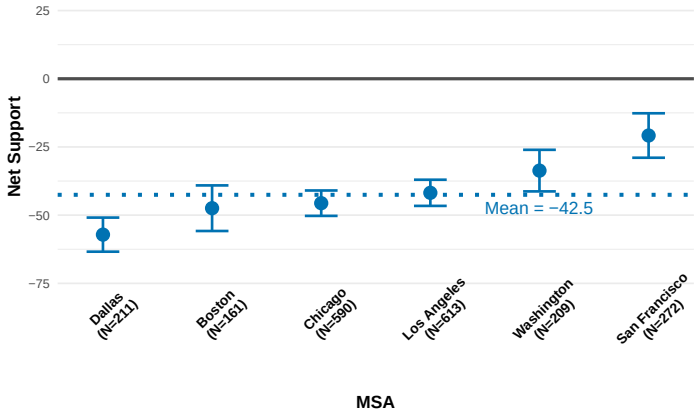
	Avg. Travel Time (min)				
Route	Pre-CP Peak	Post-CP Peak	Reduction	$\hat{\rho}$	N
Lincoln Tunnel	6.82	4.20	2.62	0.69	13,762
Holland Tunnel	15.66	10.45	5.22	0.67	13,758
Williamsburg Bridge	10.42	6.93	3.49	0.62	13,735
Queens-Midtown Tunnel	7.86	6.96	0.90	0.29	13,732
Queensboro Bridge	6.13	4.49	1.64	0.55	13,729
Mean	9.38	6.60	2.77	0.57	
Total N					68,716

Note: Congestion reduction factor $\hat{\rho}$ measures the fraction of peak-hour congestion delay eliminated after congestion pricing implementation on January 05, 2025. Data is from Moshes and Moshes (2025), and covers May 30, 2024 to April 10, 2025. Brooklyn Bridge and Manhattan Bridge are excluded due to short baseline travel times (∼2 min), which results in coarse reduction estimates, since travel time is measured in (whole) minutes. Peak hours defined as 5 AM–9 PM; uncongested baseline hours 1 AM–4 AM. $\hat{\rho}$ = (Pre-CP Peak − Post-CP Peak)/(Pre-CP Peak − Pre-CP Uncongested). N reports total readings (pre-CP + post-CP) per route.

F.2 Support by Metro Area

F.3 Treatment Effect with Sampling Weights

Figure F.1: Net Support by Metropolitan Statistical Level



Note: Net support across the six Metropolitan Statistical Areas (MSAs) in the sample, weighted by sampling weight. Net support is defined as (%support − %oppose), where each category includes both "strongly" and "somewhat." Error bars: standard errors.

Table F.2: Treatment Effects on Net Support (Sampling Weights Applied)

	Downtown Workers			Whole Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
EN-robust	41.242	39.419	43.048	7.176	7.328	8.369
	(11.965)	(12.880)	(14.491)	(3.401)	(3.411)	(3.499)
Toll Reduction	1.493	2.988	0.587	-12.176	-11.576	-9.043
	(14.013)	(14.600)	(15.748)	(4.031)	(4.042)	(4.021)
Baseline Support	✓	✓	✓	✓	✓	✓
Basic Demographics		✓	✓		✓	✓
All Demographics			✓			✓
Sampling Weights	✓	✓	✓	✓	✓	✓
R^2	0.410	0.438	0.688	0.347	0.352	0.399
Observations	184	184	181	2043	2043	2021

Note: The table reports regressions of net support for the policy on treatment, controlling for baseline support; the omitted category is Control. Columns 1–3 restrict the sample to respondents who work in the downtown area, and are eligible for contingent compensation. Columns 4–6 include the whole sample. "Basic demographics" control for income, sex, age, and education. "All demographics" additionally include ethnicity, metropolitan area, housing type, employment status, marital status, partisanship, ideology, trust in media, religion and religious attendance, household size, and phone ownership. Observations are weighted by their sampling weights. Standard errors are in parentheses.

F.4 Treatment Effect by Duration

Table F.3: Treatment Effects on Net Support, by Duration

	By Duration		
	(1)	(2)	(3)
A. Main Effects			
EN-robust	-1.423	-1.590	-0.382
	(5.139)	(5.142)	(5.170)
Toll Reduction	-9.877	-9.252	-8.010
	(5.100)	(5.103)	(5.118)
B. Interaction Effects			
EN-robust x Above Median Duration	18.176	17.669	17.238
	(6.937)	(6.958)	(6.991)
Toll Reduction x Above Median Duration	2.310	2.215	3.502
	(8.416)	(8.411)	(8.430)
Baseline Support	✓	✓	✓
Basic Demographics		✓	✓
All Demographics			✓
R^2	0.378	0.380	0.406
Observations	2043	2043	2043

Note: The table reports regressions of net support for the policy on treatment, controlling for baseline support; the omitted category is Control. Treatment is interacted with an indicator for whether the respondent's response time on the second-stage support question was above or below the median of 31 seconds. "Basic demographics" control for income, sex, age, and education. "All demographics" additionally include ethnicity, metropolitan area, housing type, homeownership, employment status, marital status, partisanship, ideology, trust in media, religion and religious attendance, household size, and phone ownership. Standard errors are in parentheses.

F.5 Treatment Effect on Beliefs

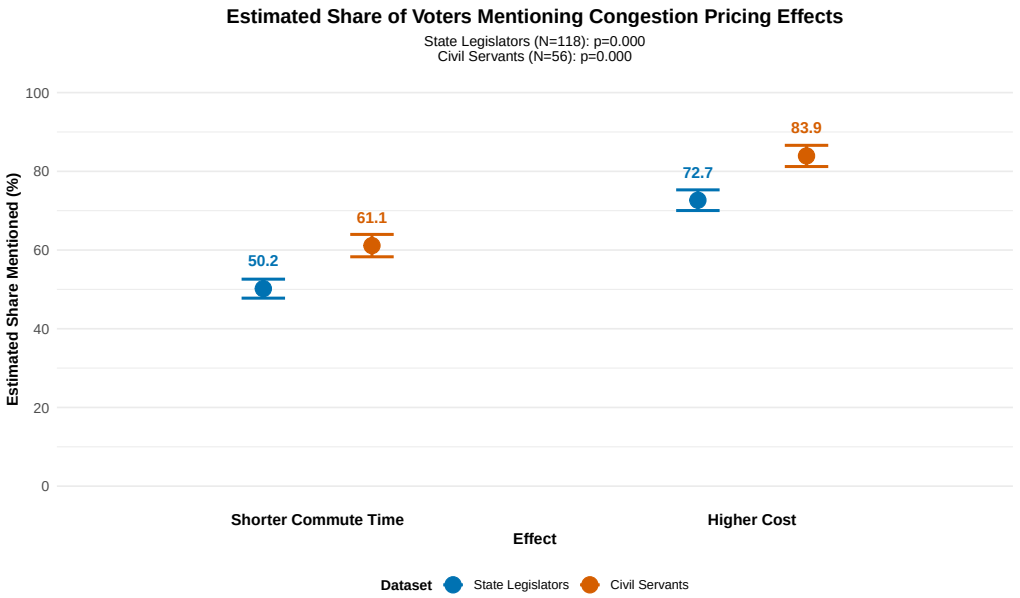
Table F.4: Treatment Effects on Beliefs

	Own Commute	NYC Commute
	(1)	(2)
EN-robust	−0.123	−0.621
	(0.608)	(0.591)
Toll Reduction	0.533	−0.468
	(0.724)	(0.71)
Baseline Support	✓	✓
R^2	0.009	0.009
Observations	977	1036

Note: The table reports regressions of second-stage beliefs in traffic reduction on treatment, controlling for baseline support; the omitted category is Control. Column 1 uses beliefs about the respondent's own commute time as the dependent variable. Column 2 uses beliefs about traffic reduction in NYC. Standard errors are in parentheses.

G Policymakers: Additional Results

Figure G.1: Policymakers: Second-order beliefs



Note: Policymakers' mean estimates of the share of the general public mentioning shorter commute times and cost increase as a result of congestion pricing. Error bars indicate standard errors. Orange: civil servants ($N = 56$). Blue: state legislators ($N = 118$). p-values: paired t-tests of the null of equal means.

Table G.1: Net Support and First- and Second-order Beliefs

	State Legislators		Civil Servants	
	(1)	(2)	(3)	(4)
Time savings (% public)	0.931	0.772	2.033	1.044
	(0.283)	(0.329)	(0.420)	(0.429)
Time savings (own)		16.537		107.353
		(17.302)		(24.597)
R^2	0.085	0.092	0.302	0.487
Observations	118	118	56	56

Note: Net support for congestion pricing on first- and second-order beliefs. Time savings (% public) is a respondent's estimate of the share of the general public mentioning time savings when asked to predict the consequences. Time savings (own) is an indicator that equals one when a respondent predicts any time savings (as opposed to an increase or "no meaningful change").